(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 January 2009 (29.01.2009)

PCT

(10) International Publication Number
WO 2009/015299 A1

(51) International Patent Classification:
C12M 1/40 (2006.01) *C12Q 1/58* (2006.01)

(21) International Application Number:
PCT/US2008/071090

(22) International Filing Date: 25 July 2008 (25.07.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/952,194 26 July 2007 (26.07.2007) US

(71) Applicant *(for all designated States except US)*: **CALIFORNIA PACIFIC MEDICAL CENTER** [US/US]; 475 Brannan Street, Suite 220, San Francisco, CA 94107 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **LEE, Nancy, M.** [US/US]; 1830 Funston Avenue, San Francisco, CA 94116 (US).

(74) Agent: **BAKER, Joseph, R. Jr.**; Gavrilovich, Dodd & Lindsey LLP, 4445 Eastgate Mall, Suite 200, San Diego, CA 92121 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: METHOD TO PREDICT OR DIAGNOSE A GASTROINTESTINAL DISORDER OR DISEASE

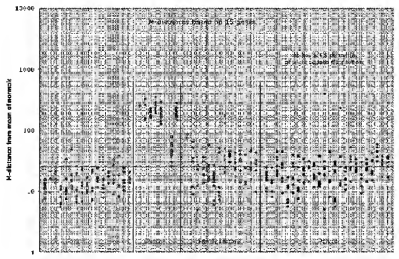

FIGURE 1A

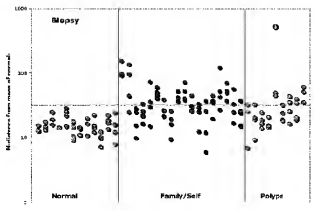

FIGURE 1B

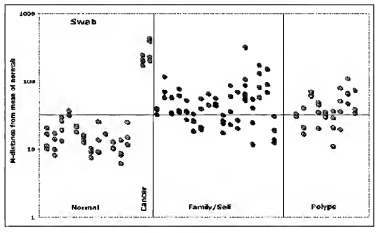

FIGURE 1C

(57) Abstract: The disclosure provides methods and compositions useful for identifying a subject's predisposition to a gastrointestinal disease or disorder.

WO 2009/015299 A1

Published:

— *with international search report*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

METHOD TO PREDICT OR DIAGNOSE A GASTROINTESTINAL DISORDER OR DISEASE

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] The application claims priority under 35 U.S.C. §119 to U.S. Provisional Application Serial No. 60/952,194, filed July 26, 2007, the disclosure of which is incorporated herein by reference.

TECHNICAL FIELD

[0002] The invention relates to predicting the probability that a subject has a predisposition to or has a gastrointestinal tract disease or disorder.

BACKGROUND

[0003] Presently, there are no biological tests in clinical use to predict a subject's clinical development of a gastrointestinal disorder or cancer based upon gene expression profiling.

SUMMARY

[0004] The dislcosure provides a method for determining if a subject has or is at risk of having a gastrointestinal disease or disorder comprising: measuring an FHSH biomarker panel, a polyp biomarker panel, a rectal bleeding biomarker panel, a cancer biomarker panel or any combination thereof, wherein a change in one or more of the biomarker panels relative to a control is indicative of a subject that has or is at risk of having a gastrointestinal disease or disorder. In one aspect, the method comprises measuring an FHSH biomarker panel and comparing the measurements to a control wherein a change relative to the control is indicative that the subject has a predisposition or risk of developing a cancerous lesion. In another aspect, if a subject is identified as having a predisposition or risk of developing a polyp or cancerous lesion, the subject is further monitored for a polyp biomarker panel. In another aspect, the subject is further monitored for a cancer biomarker panel. In yet another aspect, a polyp or cancer biomarker panel is monitored and comparing the measurements to a control wherein a change relative to the control is indicative that the subject has or is at risk of developing a polyp or cancerous lesion.

[0005] The disclosure also provides a method of determining whether a subject has rectal bleeding comprising measuring a rectal

bleeding biomarker panel, wherein a subject that is positive for the panel has rectal bleeding.

[0006] The disclosure also provides kits and compositions for carrying out the methods described herein. In one aspect, the kit comprises a FHSH biomarker panel, a polyp biomarker panel, a cancer biomarker panel, a rectal bleeding biomarker panel or any combination thereof.

[0007] The details of one or more embodiments are set forth in the accompanying drawings and the description below. Other features, objects, and advantages will be apparent from the description and drawings, and from the claims.

BRIEF DESCRIPTION OF THE FIGURES

[0008] Figure 1A-C shows the Mahalanobis distance for biopsy samples, taken from (left to right), controls, resected colon cancer, individuals with family history, and individuals with polyps (67 subject and 15 genes), (B) shows the same analysis carried out on a second patient pool, one including individuals with no polyps or family/self history (Control), individuals with family history, individuals with polyps, and (C) shows the same analysis carried out on rectal smear samples taken from the same individuals.

[0009] Figure 2A and B shows swab data. (A) shows a 90 patient study of gene expression values for 16 genes from each subject obtained by rectal swab, controls tend to fall below the 95% chi-square distribution line. A tendency of subjects with cancer to fall above the like can be seen at the far right. (B) shows the 95% chi-square distribution of gene analysis from buccal swabs of 21 controls and 8 cancer subjects.

DETAILED DESCRIPTION

[0010] As used herein and in the appended claims, the singular forms "a," "and," and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a polynucleotide" includes a plurality of such polynucleotides and reference to "the variant" includes reference to one or more variants known to those skilled in the art, and so forth.

[0011] Also, the use of "or" means "and/or" unless stated otherwise. Similarly, "comprise," "comprises," "comprising"

"include," "includes," and "including" are interchangeable and not intended to be limiting.

[0012] It is to be further understood that where descriptions of various embodiments use the term "comprising," those skilled in the art would understand that in some specific instances, an embodiment can be alternatively described using language "consisting essentially of" or "consisting of."

[0013] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this disclosure belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice of the disclosed methods and compositions, the exemplary methods, devices and materials are described herein.

[0014] The publications discussed above and throughout the text are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the inventors are not entitled to antedate such disclosure by virtue of prior disclosure.

[0015] The disclosure provides a number of biomarkers useful for predicting a subject's predisposition or the existence of a gastrointestinal disease or disorder. The biomarkers identified herein can be used in combination with additional predictive tests including, but not limited to, additional SNPs, mutations, and clinical tests (including a plurality of biomarker panels disclosed herein).

[0016] The methods and compositions of the disclosure can be used in an outpatient clinic or inpatient environment. Outpatient clinical diagnostics are useful to reduce costs of unnecessary, often invasive or painful, procedures. As a screening tool, colonoscopy is considered too expensive, both to the patients and to the insurance carriers, and carries with it a small percentage of risks and complications. Barium enema and CT colonography (or virtual colonoscopy), like colonoscopy, will provide for a complete colon examination, but small polyps or even small cancers can be missed. The cost is high, and higher still if a polyp or cancer or even a suggestion of a polyp or cancer will be interpreted by the

radiologists, requiring the additional procedure of colonoscopy for confirmation. The barium enema, the CT colonography and the colonoscopy procedures all require the patients to have a thorough mechanical bowel preparation the day before. The diagnostic tests and compositions described herein are useful to identify, diagnose, and prognose subjects that should be followed or treated for gastrointestinal diseases and disorders including the development of polyps, cancerous lesions or other non-cancerous inflammatory diseases.

[0017] An adenoma, colon adenoma, and polyp are used herein to describe any precancerous neoplasia of the colon. Precancerous colon neoplasias are referred to as adenomas or adenomatous polyps. Adenomas are typically small mushroom-like or wart-like growths on the lining of the colon and do not invade into the wall of the colon. Adenomas may be visualized through a device such as a colonoscope or flexible sigmoidoscope. Several studies have shown that patients who undergo screening for and removal of adenomas have a decreased rate of mortality from colon cancer. For this and other reasons, it is generally accepted that adenomas are an obligate precursor for the vast majority of colon cancers. When a colon neoplasia invades into the basement membrane of the colon, it is considered a colon cancer. The most widely used staging systems generally use at least one of the following characteristics for staging: the extent of tumor penetration into the colon wall, with greater penetration generally correlating with a more dangerous tumor; the extent of invasion of the tumor through the colon wall and into other neighboring tissues, with greater invasion generally correlating with a more dangerous tumor; the extent of invasion of the tumor into the regional lymph nodes, with greater invasion generally correlating with a more dangerous tumor; and the extent of metastatic invasion into more distant tissues, such as the liver, with greater metastatic invasion generally correlating with a more dangerous disease state.

[0018] An allele refers to a particular form of a genetic locus, distinguished from other forms by its particular nucleotide sequence, or one of the alternative polymorphisms found at a polymorphic site.

[0019] A biological sample refers to a sample obtained from a subject wherein the sample comprises cells, or can be cell free. The biological sample can be blood, sputum, saliva, tissue, stool, urine, serum cerebrospinal, cells, secretions or the like. Where the sample is a tissue, the tissue sample can be obtained by biopsy. Biopsy samples can be obtained from the gastrointestinal tract (*e.g.*, from a segment of colon between the cecum and the hepatic flexure were classified as ascending colon samples; those from the segment of colon between the hepatic flexure and the splenic flexure as transverse colon samples; those from the segment of colon below the splenic flexure as descending colon; those from the winding segment of colon below the descending colon were classified as rectosigmoid colon samples (approximately 5-25 cm from rectum, typically about 5-10 cm)). The biological sample can be obtained non-invasively (*e.g.*, by swab). The swab, for example, can be obtained from the mouth or rectum. In one embodiment, the swab is obtained from the distal portion of the gastrointestinal tract (*e.g.*, the last 5-25 cm is obtained from the rectum). In yet another embodiment, the swab is collected from the buccal area (e.g., the mouth, cheek, sublingual area, gums and the like). A minimally invasive method, such as a swab, or a non-invasive sampling method, such as a stool sample can be obtained and used in the methods of the disclosure. A biopsy will tend to have a more heterogenous mixture of cell-types (*e.g.*, epithelial, stromal and endothelial cells) compared to a swab sample, which has a higher percentage of cell types on the colorectal surface (*e.g.*, epithelial and inflammatory cells).

[0020] A biomarker refers to a detectable biological entity associated with a particular phenotype or risk of developing a particular phenotype. The biological entity can be a polypeptide or polynucleotide. A biomarker to be detected is referred to as a target. For example, a target polynucleotide refers to a biomarker comprising a polynucleotide (*e.g.*, an mRNA or cDNA) that is to be detected. In another example, a target polypeptide refers to a protein expressed (*i.e.*, transcribed and translated) that is to be detected. A biomarker, as defined by the National Institutes of Health (NIH), refers to a molecular indicator of a specific

biological property; a biochemical feature or facet that can be used to measure the progress of disease or the effects of treatment. A panel of biomarkers is a selection of at least two biomarkers. Biomarkers may be from a variety of classes of molecules. In principle, the larger the number of biomarkers used the more sensitive the analysis will be. The panel can comprise from 2 to sixteen or more biomarkers. In one aspect, the panel comprises 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 or more biomarkers. The disclosure demonstrates that for individuals with cancer, three or four genes, such as COX-2, IL-8 and CD44, can suffice. However, for individuals with polyps or with history of cancer fine-tuning the analysis by adding to or otherwise modifying the biomarker panel increases specificity.

[0021] The term "colon" as used herein is intended to encompass the right colon (including the cecum), the transverse colon, the left colon, and the rectum.

[0022] A colorectal cancer and colon cancer are used interchangeably herein to refer to any cancerous neoplasia of the colon (including the rectum). The concept of polyp to cancer sequence is well established, and it is widely accepted that removal of pre-malignant colorectal polyps will lead to a significant decrease of the incidence of colorectal cancer. Furthermore, clinical data has showns that early detection and curative surgical resection of colorectal cancer will significantly improve survival rates.

[0023] Subjects with either a family history of any cancer or personal history of any cancer and with no polyps during a current colonoscopy are referred to as FHSH subjects. Subjects with polyps and with or without family or self history of any cancer are referred to as polyps subjects and comprise a FHSH subject's biomarker panel. Subjects with colon cancer are referred to as cancer subjects and comprise a cancer subject's biomarker panel.

[0024] A fecal occult blood test (FOBT) is a test used to check for hidden blood in the stool. Sometimes cancers or polyps can bleed, and FOBT is used to detect small amounts of bleeding. In addition, screening tests (such as a rectal examination, proctoscopy, and colonoscopy) may be done regularly in patients who

are at high risk of colon cancer or who have a positive FOBT and/or biomarker results. The proctoscopy examination finds about half of all colon and rectal cancers. After treatment, a blood test and x-rays may be done to screen for recurrence.

[0025] Colorectal cancer, also referred to as colon cancer or large bowel cancer, includes cancerous growths in the colon, rectum and appendix. Many colorectal cancers arise from adenomatous polyps in the colon. These growths are usually benign, but some may develop into cancer over time. The majority of the time, the diagnosis of localized colon cancer is through colonoscopy. Therapy is usually through surgery, which in many cases is followed by chemotherapy. Polyps of the colon, particularly adenomatous polyps, are a risk factor for colon cancer. The removal of colon polyps at the time of colonoscopy reduces the subsequent risk of colon cancer. Individuals who have previously been diagnosed and treated for colon cancer are at risk for developing colon cancer in the future. Women who have had cancer of the ovary, uterus, or breast are at higher risk of developing colorectal cancer. Family history of colon cancer, especially in a close relative before the age of 55 or multiple relatives, increases the risk of cancer in a subject.

[0026] Gastrointestinal inflammation refers to inflammation of a mucosal layer of the gastrointestinal tract, and encompasses acute and chronic inflammatory conditions. Acute inflammation is generally characterized by a short time of onset and infiltration or influx of neutrophils. Chronic inflammation is generally characterized by a relatively longer period of onset and infiltration or influx of mononuclear cells. Chronic inflammation can also be characterized by periods of spontaneous remission and spontaneous occurrence. The mucosal layer of the gastrointestinal tract includes mucosa of the bowel (including the small intestine and large intestine), rectum, stomach (gastric) lining, oral cavity, and the like.

[0027] Chronic gastrointestinal inflammation refers to inflammation of the mucosa of the gastrointestinal tract that is characterized by a relatively longer period of onset, is long-lasting (*e.g.*, from several days, weeks, months, or years and up to

the life of the subject), and is associated with infiltration or influx of mononuclear cells and can be further associated with periods of spontaneous remission and spontaneous occurrence. Examples of chronic gastrointestinal inflammation include inflammatory bowel disease (IBD), colitis induced by environmental insults (*e.g.*, gastrointestinal inflammation (*e.g.*, colitis) caused by or associated with (*e.g.*, as a side effect) a therapeutic regimen, such as administration of chemotherapy, radiation therapy, and the like), colitis in conditions such as chronic granulomatous disease (Schappi *et al.* Arch Dis Child. 2001 Feb;84(2):147-151), celiac disease, celiac sprue (a heritable disease in which the intestinal lining is inflamed in response to the ingestion of a protein known as gluten), food allergies, gastritis, infectious gastritis or enterocolitis (*e.g.*, Helicobacter pylori-infected chronic active gastritis) and other forms of gastrointestinal inflammation caused by an infectious agent, and other like conditions.

[0028] As used herein, "inflammatory bowel disease" or "IBD" refers to any of a variety of diseases characterized by inflammation of all or part of the intestines. Examples of inflammatory bowel disease include, but are not limited to, Crohn's disease, Barrett's disease and ulcerative colitis. Reference to IBD throughout the specification is often referred to in the specification as exemplary of gastrointestinal inflammatory conditions, and is not meant to be limiting. The term IBD includes pseudomembranous colitis, hemorrhagic colitis, hemolytic-uremic syndrome colitis, collagenous colitis, ischemic colitis, radiation colitis, drug and chemically induced colitis, diversion colitis, ulcerative colitis, irritable bowel syndrome, irritable colon syndrome, Barrett's disease and Crohn's disease; and within Crohn's disease all the subtypes including active, refractory, and fistulizing and Crohn's disease.

[0029] A non-colorectal cancer inflammatory disease or disorder of the gastrointestinal tract refers to an inflammation of the gastrointestinal tract in the absence of a cancerous lesion, tumor or lesion. A non-colorectal cancer inflammatory disease or

disorder of the gastrointestinal tract includes inflammatory bowel disease.

[0030] A gene refers to a segment of genomic DNA that contains the coding sequence for a protein, wherein the segment may include promoters, exons, introns, and other untranslated regions that control expression.

[0031] A genotype is an unphased 5' to 3' sequence of nucleotide pair(s) found at a set of one or more polymorphic sites in a locus on a pair of homologous chromosomes in an individual. As used herein, genotype includes a full-genotype and/or a sub-genotype.

[0032] Genotyping is a process for determining a genotype of an individual.

[0033] A haplotype is a 5' to 3' sequence of nucleotides found at a set of one or more polymorphic sites in a locus on a single chromosome from a single individual.

[0034] Haplotype pair is two haplotypes found for a locus in a single individual.

[0035] Haplotyping is the process for determining one or more haplotypes in an individual and includes use of family pedigrees, molecular techniques and/or statistical inference.

[0036] A genetic locus refers to a location on a chromosome or DNA molecule corresponding to a gene or a physical or phenotypic feature, where physical features include polymorphic sites.

[0037] Polymorphic site (PS) is a position on a chromosome or DNA molecule at which at least two alternative sequences are found in a population.

[0038] A polymorphism refers to the sequence variation observed in an individual at a polymorphic site. Polymorphisms include nucleotide substitutions, insertions, deletions and microsatellites and may, but need not, result in detectable differences in gene expression or protein function. A single nucleotide polymorphism (SNP) is a single change in the nucleotide variation at a polymorphic site.

[0039] An oligonucleotide probe or a primer refers to a nucleic acid molecule of between 8 and 2000 nucleotides in length, or is about 6 and 1000 nucleotides in length. More particularly, the

length of these oligonucleotides can range from about 8, 10, 15, 20, or 30 to 100 nucleotides, but will typically be about 10 to 50 (*e.g.*, 15 to 30 nucleotides). The appropriate length for oligonucleotides in assays of the disclosure under a particular set of conditions may be empirically determined by one of skill in the art.

[0040] Oligonucleotide primers and probes can be prepared by any suitable method, including, for example, cloning and restriction of appropriate sequences and direct chemical synthesis. The oligonucleotide primers and probes can contain conventional nucleotides, as well as any of a variety of analogs. For example, the term "nucleotide", as used herein, refers to a compound comprising a nucleotide base linked to the C-1' carbon of a sugar, such as ribose, arabinose, xylose, and pyranose, and sugar analogs thereof. The term nucleotide also encompasses nucleotide analogs. The sugar may be substituted or unsubstituted. Substituted ribose sugars include, but are not limited to, those riboses in which one or more of the carbon atoms, for example the 2'-carbon atom, is substituted with one or more of the same or different Cl, F, --R, --OR, $--NR_2$ or halogen groups, where each R is independently H, C_1-C_6 alkyl or C_5-C_{14} aryl. Exemplary riboses include, but are not limited to, 2'-(C_1-C_6)alkoxyribose, 2'-(C_5-C_{14})aryloxyribose, 2',3'-didehydroribose, 2'-deoxy-3'-haloribose, 2'-deoxy-3'-fluororibose, 2'-deoxy-3'-chlororibose, 2'-deoxy-3'-aminoribose, 2'-deoxy-3'-(C_1-C_6)alkylribose, 2'-deoxy-3'-(C_1-C_6)alkoxyribose and 2'-deoxy-3'-(C_5-C_{14})aryloxyribose, ribose, 2'-deoxyribose, 2',3'-dideoxyribose, 2'-haloribose, 2'-fluororibose, 2'-chlororibose, and 2'-alkylribose, *e.g.*, 2'-O-methyl, 4'-α-anomeric nucleotides, 1'-α-anomeric nucleotides, 2'-4'- and 3'-4'-linked and other "locked" or "LNA", bicyclic sugar modifications (see, *e.g.*, PCT published application nos. WO 98/22489, WO 98/39352; and WO 99/14226). Exemplary LNA sugar analogs within a polynucleotide include, but are not limited to, the structures: where B is any nucleotide base.

[0041] Modifications at the 2'- or 3'-position of ribose include, but are not limited to, hydrogen, hydroxy, methoxy, ethoxy, allyloxy, isopropoxy, butoxy, isobutoxy, methoxyethyl, alkoxy, phenoxy, azido, amino, alkylamino, fluoro, chloro and

bromo. Nucleotides include, but are not limited to, the natural D optical isomer, as well as the L optical isomer forms (see, *e.g.*, Garbesi (1993) Nucl. Acids Res. 21:4159-65; Fujimori (1990) J. Amer. Chem. Soc. 112:7435; Urata, (1993) Nucleic Acids Symposium Ser. No. 29:69-70). When the nucleotide base is purine, *e.g.* A or G, the ribose sugar is attached to the N_9-position of the nucleotide base. When the nucleotide base is pyrimidine, *e.g.* C, T or U, the pentose sugar is attached to the N_1-position of the nucleotide base, except for pseudouridines, in which the pentose sugar is attached to the C_5 position of the uracil nucleotide base (see, *e.g.*, Kornberg and Baker, (1992) DNA Replication, 2nd Ed., Freeman, San Francisco, Calif.). The 3' end of the probe can be functionalized with a capture or detectable label to assist in detection of a target polynucleotide or of a polymorphism.

[0042] Any of the oligonucleotides or nucleic acids of the disclosure can be labeled by incorporating a detectable label measurable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, such labels can comprise radioactive substances (*e.g.*, ^{32}P, ^{35}S, ^{3}H, ^{125}I), fluorescent dyes (*e.g.*, 5-bromodesoxyuridin, fluorescein, acetylaminofluorene, digoxigenin), biotin, nanoparticles, and the like. Such oligonucleotides are typically labeled at their 3' and 5' ends.

[0043] A probe refers to a molecule which can detectably distinguish changes in gene expression or can distinguish between target molecules differing in structure. Detection can be accomplished in a variety of different ways depending on the type of probe used and the type of target molecule. Thus, for example, detection may be based on discrimination of activity levels of the target molecule, but typically is based on detection of specific binding. Examples of such specific binding include antibody binding and nucleic acid probe hybridization. Thus, for example, probes can include enzyme substrates, antibodies and antibody fragments, and nucleic acid hybridization probes (including primers useful for polynucleotide amplification and/or detection). Thus, in one embodiment, the detection of the presence or absence of the at least one target polynucleotide involves contacting a biological

sample with a probe or primer pair, typically an oligonucleotide probe or primer pair, where the probe/primers hybridizes with a form of a target polynucleotide in the biological sample containing a complementary sequence, where the hybridization is carried out under selective hybridization conditions. Such an oligonucleotide probe can include one or more nucleic acid analogs, labels or other substituents or moieties so long as the base-pairing function is retained.

[0044] A reference or control population refers to a group of subjects or individuals who are predicted to be representative of the genetic variation found in the general population having a particular genotype or expression profile. Typically, the reference population represents the genetic variation in the population at a certainty level of at least 85%, typically at least 90%, least 95% and but commonly at least 99%. The reference or control population can include subjects who individually have not demonstrated any gastrointestinal disease or disorder and can include individuals whose family line does not or has not demonstrated any gastrointestinal diseases or disorders.

[0045] A subject comprises an individual (*e.g.*, a mammalian subject or human) whose gene expression profile, genotypes or haplotypes or response to treatment or disease state are to be determined. A control subject refers to individuals with no polyps and no family or self history of cancer or known upper GI problem. Subjects with either a family history of any cancer or personal history of any cancer, and with no polyps during a current colonoscopy are referred to as FHSH subjects. Subjects with polyps and with or without family or self history of any cancer are referred to as polyps subjects and comprise a FHSH subject's biomarker panel.

[0046] In some instances a subject may not have access or know their familial history. In such instances, the diagnostics of the disclosure can be used to determine if they have a predisposition to a gastrointestinal disease or disorder based upon a FHSH biomarker panel. In other aspects, where a subject is identified as having a FHSH GI disease or disorder, the subject may be monitored for changes in biomarker expression indicative of cancer

lesions or polyps based upon a cancer biomarker panel. Where a biomarker panel associated with colorectal cancer is present the subject may be monitored by, for example, by colonoscopy for early detection and removal of polyps or cancerous lesions. One advantage of the biomarker panels provided herein is that the panel may be detected by swab collection (*e.g.*, swab of the rectal 5-10 cm) or a buccal swab. Such procedures may be performed in an outpatient setting. As indicated above, statistics indicate that early detection and removal of cancerous lesion and polyps reduce mobidity and mortality of subjects.

[0047] One embodiment of what is disclosed is the measurement of at least one or a panel of biomarkers with the selectivity and sensitivity required for managing and diagnosing subjects that have or may have a predisposition to a gastrointestinal disease or disorder. Table 1 provides a list of polynucleotide biomarkers useful in the methods and compositions of the disclosure (each of the sequences associated with the Enterez Accession Nos. set forth in Table 1 are incorporated herein by reference).

TABLE 1

SEQ ID NO: polynucleotide and polypeptide	**NCBI Entrez Database**	**Name**	**Abbreviation**
1 and 2	XM_031289	Interleukin-8	IL8
3 and 4	NM_000389	cyclin-dependent kinase inhibitor 1A (p21, Cip1)	P21
5 and 6	XM_030326	CD44 antigen	CD44
7 and 8	M94582	Interleukin 8 receptor B	CXCR2
9 and 10	X54489	Melanoma growth stimulatory activity	Gro-alpha
11 and 12	NM_002090	Chemokine (C-X-C motif) ligand3	Gro-gamma
13 and 14	XM_003059	Peroxisome proliferative activated receptor, gamma	PPAR-gamma
15 and 16	NM_006238	Peroxisome proliferative activated receptor, delta	PPAR-delta
17 and 18	AX057136	c-Myc	c-Myc
19 and 20	XM_032429	Secreted phosphoprotein 1	SPP1 (OPN)
21 and 22	XM_044882	Prostaglandin-endoperoxide synthase 1	COX-1

23 and 24	XM_051900	Prostaglandin-endoperoxide synthase 2	COX-2
25 and 26	NM_005036	Peroxisome proliferative activated receptor, alpha	PPAR-alpha
27 and 28	NM_000757	Macrophage colony stimulating factor 1	MCSF-1
29 and 30	M64349	Cyclin-D	Cyc-D
31 and 32	NM_000331	Serum amyloid A1	SAA1
33 and 34	NM_002131	Homo sapiens high mobility group AT-hook 1 (HMGA1)	HMGA1
35 and 36	X54942 X55506	CKSHS2	CKSHS2
37 and 38	U22055	Human 100 kDa coactivator	p100 activator
39 and 40	NM_005555	Homo sapiens keratin 6B	LCN2
41 and 42	BC021998	Homo sapiens cyclin-dependent kinase inhibitor 2A	hCDK2a
43 and 44	NM_058195	Homo sapiens cyclin-dependent kinase inhibitor 2A	hCDK2a alt.

[0048] Naturally occurring variants (*e.g.*, polymorphisms) of any of the foregoing polynucleotides identified in Table 1 are encompassed by the disclosure. Identification of such naturally occurring polymorphisms are routinely identified or are known in the art. For example, polymorphisms of IL-8 and CXCR2 include SNP -251, -353/+1530, -353/+3331, and +1530/+3331 of IL-8 and +785/+1208 of CXCR2. Others include IL1B -31 SNP (C to T), IL10 -819 T/T. RS numbers include rs1143627 (IL1B), rs2243250 and rs1143634 (IL4), rs1801282 (PPAR-gamma), rs4073 (IL8), rs1800629 (TNF), and rs20417, rs5277, rs20432 and rs5275 (COX2).

[0049] In one aspect of the disclosure, expression levels of polynucleotides comprising biomarkers, or fragments thereof, indicated in SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, 17 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41 or 43 are used in the determination of a gastrointestinal disease or disorder or a predisposition to a gastrointestinal disease or disorder. Such analysis of polynucleotide expression levels is frequently referred to in the art as gene expression profiling. In gene expression profiling, levels of mRNA in a sample are measured as a leading

indicator of a biological state, in this case, as an indicator of a gastrointestinal disease or disorder or a predisposition thereto. One of the most common methods for analyzing gene expression profiling is to create multiple copies from mRNA in a biological sample using a process known as reverse transcription. In the process of reverse transcription, the mRNA from the sample is used to create DNA copies of the corresponding mRNA. The copies made from mRNA are referred to as copy DNA, or cDNA. mRNA is somewhat unstable and subject degradation by RNAses. In one aspect, the RNA can be protected by using RNAse inhibitors and cocktails known in the art. Table 2 provides probes and primers useful to detecting a polynucleotide biomarker of the disclosure.

Table 2

Sequence ID No./ID	Sequence	Name
45. Forward Primer	agatattgca cgggagaata tacaaa	Interleukin 8
46. Reverse Primer	tcaattcctg aaattaaagt tcggata	
47. Forward Primer	tctgcagagt tggaagcact cta	Prostaglandin-endoperoxide synthase 2
48. Reverse Primer	gccgaggctt ttctaccaga a	
49. Forward Primer	catggcttga tcagcaagga	Interleukin 8 receptor B (CXCR2)
50. Reverse Primer	tggaagtgtg ccctgaagaa g	
51. Forward Primer	caaggagctg acttcggaac taa	Lipocalin 2
52. Reverse Primer	agggaagacg atgtggtttt ca	
53. Forward Primer	gggacatgtg gagagcctac tc	Serum amyloid A1
54. Reverse Primer	catcatagtt cccccgagca t	
55. Forward Primer	aagcagcacc agcaagtgaa g	Macrophage colony stimulating factor 1
56. Reverse Primer	tcatggcctg tgtcagtcaa a	
57. Forward Primer	acatgccagc cactgtgata g	Melanoma growth stimulatory activity
58. Reverse Primer	ccctgccttc acaatgatct c	
59. Forward Primer	ggaattcacc tcaagaacat cca	Chemokine (C-X-C motif) ligand 3
60. Reverse Primer	agtgtggcta tgacttcggt ttg	
61. Forward Primer	cagccacaag cagtccagat ta	(OPN) Secreted phosphoprotein 1
62. Reverse Primer	cctgactatc aatcacatcg gaat	
63. Forward Primer	ccaggtgctc cacatgacag t	Cyclin D
64. Reverse Primer	aaacaaccaa caacaaggag aatg	
65. Forward Primer	cgtctccaca catcagcaca a	c-Myc
66. Reverse Primer	tcttggcagc aggatagtcc tt	
67. Forward Primer	gcagaccagc atgacagatt tc	Cyclin-dependent kinase inhibitor (p21)
68. Reverse Primer	gcggattagg gcttcctctt	
69. Forward Primer	ggcaccagag gcagtaacca t	Cyclin-dependent kinase inhibitor 2A

Sequence ID No./ID	Sequence	Name
70. Reverse Primer	agcctctctg gttctttcaa tcg	
71. Forward Primer	tggttcacat cccgcggct	Alternative reading frame p14
72. Reverse Primer	tggctcctca gtagcatcag	
73. Forward Primer	tgaagttcaa tgcactggaa ctg	Peroxisome proliferation activated receptor, alpha
74. Reverse Primer	caggacgatc tccacagcaa	
75. Forward Primer	tggagtccac gagatcattt aca	Peroxisome proliferation activated receptor, gamma
76. Reverse Primer	agccttggcc ctcggatat	
77. Forward Primer	cactgagttc gccaagagca t	Peroxisome proliferation activated receptor, delta
78. Reverse Primer	cacgccatac ttgagaaggg taa	
79. Forward Primer	gctagtgatc aacagtggca atg	CD44 antigen
80. Reverse Primer	gctggcctct ccgttgag	
81. Forward Primer	tgttcggtgt ccagttccaa ta	Prostaglandin-endoperoxide synthase 1
82. Reverse Primer	tgccagtggt agagatggtt ga	
83. Forward Primer	acaactccag gaaggaaacc aa	High-mobility group AT-hook1 isoform B
84. Reverse Primer	cgaggactcc tgcgagatg	
85. Forward Primer	tgaagaggag tggaggagac ttg	CKS1 protein homolog
86. Reverse Primer	gaatatgtgg ttctggctca tgaa	
87. Forward Primer	gagaaggagc gatctgctag ct	100 kDa coactivator
88. Reverse Primer	cacgtagaag tgcaggtcat cag	

[0050] Methods known in the art can be used to quantitatively measure the amount of mRNA transcribed by cells present in a sample. Examples of such methods include quantitative polymerase chain reaction (PCR), digital PCR, northern and southern blots. PCR allows for the detection and measurement of very low quantities of mRNA using an amplification process. Genes may either be up regulated or down regulated in any particular biological state, and hence mRNA levels shift accordingly.

[0051] The following tables identify various biomarker panels and statistics useful in performing the diagnostics of the disclosure.

[0052] A polyp biomarker panel based upon a swab comprises one or more of the biomarkers CD44, PPARγ, and COX1. In one aspect, a polyp biomarker panel using a swab comprises the genes listed in Table 3. The percentage shown in Tables 3-10 comprises the percentage of subject in the population showing a change (e.g., an

increase or decrease in expression) in the listed biomarkers compared to a control population.

Table 3

Swabs Polyps	% having a change relative to control
CD44	45.5% ± 2.5%
PPARγ	40.5% ± 2.5%
COX 1	45.5% ± 2.5%
PPARα	37.0% ± 1.0%
SAA1	38.0% ± 1.0%
OPN=COX2=IL8=cMyc=mCSF1=cycD	31.0% ± 2.0%
Groα	29.0% ± 1.0%
PPARδ	18.0% ± 5.0%
P21=Groγ	19.0% ± 1.0%

[0053] A polyp biomarker panel based upon a rectal biopsy comprises one or more of the biomarkers Groα, CXCR2, and PPARδ. The biomarker panel can further comprise P21. In one aspect, a rectal polyp biomarker panel using a biopsy comprises the genes listed in Table 4.

TABLE 4

Rectal Biopsy Polyps	% having a change relative to control
Groα	60.0% ± 1.0%
CXCR2	55.0% ± 1.0%
PPARδ	45.0% ± 1.0%
P21	30.0% ± 1.0%
OPN=PPARα=CD44	25.0% ± 1.0%
PPARγ=SAA1=COX1	20.0% ± 1.0%
Groγ=cMyc=mCSF1	15.0% ± 1.0%
cycD	5.0% ± 1.0%
COX2	0%

[0054] A polyp biomarker panel based upon an ascending colon biopsy comprises one or more of the biomarkers P21, mCSF-1, cycD, and SAA1. In one aspect, an ascending colon polyp biomarker panel using a biopsy comprises the genes listed in Table 5.

TABLE 5

AS Biopsy Polyps	% having a change relative to control
P21=mCSF1	45.0% ± 1.0%
cycD	41.0% ± 1.0%
SAA1	32.0% ± 1.0%
Groα=OPN=CXCR2=PPARα=CD44	27.0% ± 1.0%
COX 1= Groγ=IL-8	23.0% ± 1.0%

PPARδ	18.0% ± 1.0%
COX2	14.0% ± 1.0%
cMyc=PPARγ	5.0% ± 1.0%

[0055] A polyp biomarker panel based upon a descending colon biopsy comprises one or more of the biomarkers COX-1, CXCR2, cycD, PPARδ and SAA1. In one aspect, a descending colon polyp biomarker panel using a biopsy comprises the genes listed in Table 6.

TABLE 6

DS Biopsy Polyps	**% having a change relative to control**
CXCR2=COX1	39.0% ± 1.0%
cycD=PPARδ	35.0% ± 1.0%
SAA1	30.0% ± 1.0%
PPARγ=P21	26.0% ± 1.0%
mCSF-1=cMyc=Groα	22.0% ± 1.0%
CD44=PPARα	17.0% ± 1.0%
IL-8=COX2	13.0% ± 1.0%
OPN=Groγ	9.0% ± 1.0%

[0056] A FHSH biomarker panel based upon a rectal swab comprises one or more of the biomarkers Groα, CD44, and COX1. In one aspect, a FHSH biomarker panel using a swab comprises the genes listed in Table 7.

TABLE 7

SWABS FHSH	**% having a change relative to control**
Groα	50.0% ± 1.0%
CD44	46.0% ± 1.0%
COX1=Groγ	42.0% ± 1.0%
OPN=COX2=cMyc	38.0% ± 1.0%
mCSF-1	33.0% ± 2.0%
PPARγ=P21=cycD=PPARδ	31.0% ± 1.0%
SAA1	27.0% ± 1.0%
IL8	23.0% ± 1.0%
CXCR2	19.0% ± 1.0%
PPARα	15.0% ± 1.0%

[0057] A FHSH biomarker panel based upon a rectal biopsy comprises one or more of the biomarkers GROα, PPARδ, SAA1, COX1 and CXCR2. In one aspect, a rectal biopsy FHSH biomarker panel using a biopsy comprises the genes listed in Table 8.

TABLE 8

RECTAL BIOPSIES FHSF	% having a change relative to control
Groα=PPARδ=SAA1	40.0% ± 1.0%
COX1=CXCR2	36.0% ± 1.0%
cMyc=CD44	32.0% ± 1.0%
P21	28.0% ± 1.0%
OPN=PPARα=COX2	24.0% ± 1.0%
Groγ	20.0% ± 1.0%
IL8	16.0% ± 1.0%
PPARγ=mCSF1	12.0% ± 1.0%
cycD	4.0% ± 1.0%

[0058] A FHSH biomarker panel based upon an ascending colon biopsy comprises one or more of the biomarkers m-CSF1, p21, and cycD. In one aspect, a ascending colon biopsy FHSH biomarker panel using a biopsy comprises the genes listed in Table 9.

TABLE 9

AS BIOPSIES FHSF	% having a change relative to control
mCSF1	60.0% ± 1.0%
P21	46.0% ± 1.0%
cycD	40.0% ± 1.0%
SAA1=cMyc=CXCR2=Groγ	26.0% ± 1.0%
Groα=IL8=Cox1	23.0% ± 1.0%
CD44	20.0% ± 1.0%
PPARδ	17.0% ± 1.0%
OPN	14.0% ± 1.0%
PPARα=COX-2=PPARγ	11.0% ± 1.0%

[0059] A FHSH biomarker panel based upon a descending colon biopsy comprises one or more of the biomarkers CXCR2, cycD and SAA1. In one aspect, a descending colon biopsy FHSH biomarker panel using a biopsy comprises the genes listed in Table 10.

TABLE 10

DS BIOPSIES FHSF	% having a change relative to control
CXCR2	42.0% ± 1.0%
cycD	39.0% ± 1.0%
SAA1	33.0% ± 1.0%
mCSF1-PPARδ	31.0% ± 1.0%
Groγ	28.0% ± 1.0%
P21=COX2=Groα	25.0% ± 1.0%
PPARγ	19.0% ± 1.0%

cMyc=IL8	17.0% ± 1.0%
CD44=OPN	11.0% ± 1.0%
PPARα=COX1	8.0% ± 1.0%

[0060] A rectal bleeding biomarker panel based upon a swab comprises one or more of the biomarkers COX2, OPN, PPARγ, COX1 and GROα. In one aspect, a rectal bleeding biomarker panel using a swab comprises the genes listed in Table 11. Rectal bleeding biomarkers can be indicative of a non-cancerous inflammatory disease or disorder.

TABLE 11

SWABS RECTAL BLEEDING	**% having a change relative to control**
COX2	53.0% ± 1.0%
OPN=PPARγ	47.0% ± 1.0%
COX1=Groα	40.0% ± 1.0%
CXCRZ=IL8=CD44=cycD	33.0% ± 1.0%
PPARα=Groγ=PPARδ	27.0% ± 1.0%
P21	20.0% ± 1.0%
cMyc=mCSF1	13.0% ± 1.0%
SAA1	7.0% ± 1.0%

[0061] A rectal bleeding biomarker panel based upon a biopsy comprises one or more of the biomarkers Groα, Groγ, PPARδ and SAA1. In one aspect, a rectal bleeding biomarker panel using a biopsy comprises the genes listed in Table 12.

TABLE 12

BIOPSIES RECTAL BLEEDING	**% having a change relative to control**
Groα=Groγ=PPARδ	54.0% ± 1.0%
SAA1	46.0% ± 1.0%
CXCR2=mCSF1	38.0% ± 1.0%
OPN=PPARα=CD44	31.0% ± 1.0%
COX2=cMyc	23.0% ± 1.0%
IL8=PPARγ=P21=cycD	15.0% ± 1.0%
COX1	13.0% ± 1.0%

[0062] A cancer biomarker panel based upon a swab in the absence of an RNA protection cocktail comprises the biomarkers PPARα, CXCR2, cMyc and CD44. In one aspect, a cancer biomarker panel using a swab comprises the genes listed in Table 13.

TABLE 13

SWABS CANCER (PBS)	**% having a change relative to**

	control
CXCR2=PPARα=cMyc=CD44	100%
OPN=COX1=COX2=Groα=Groγ=IL8=PPARγ=P21=SAA1	75.0% ± 1.0%
cycD=PPARδ	50.0% ± 1.0%
mCSF1	0%

[0063] A cancer biomarker panel based upon a swab in the presence of an RNA protection cocktail comprises the biomarkers COX2 and IL-8. In one aspect, a cancer biomarker panel using a swab comprises the genes listed in Table 14.

TABLE 14

SWABS CANCER (RNA PROTECTION)	% having a change relative to control
COX2=IL8	100%
Groγ=COX1=CD44	67.0% ± 1.0%
OPN=cMyc=mCSF1=cycD	50.0% ± 1.0%
CXCRZ=Groα=PPARγ=P21	33.0% ± 1.0%
PPARα=PPARδ	17.0% ± 1.0%
SAA1	0%

[0064] In one embodiment, a method for gene expression profiling comprises measuring mRNA levels for biomarkers selected in a panel. Such a method can include the use of primers, probes, enzymes, and other reagents for the preparation, detection, and quantitation of mRNA (*e.g.*, by PCR, by Northern blot and the like). The primers listed in SEQ ID NOs: 45-88 are particularly suited for use in gene expression profiling using RT-PCR based on a polynucleotide biomarker. Although the disclosure provides particular primers and probes, those of skill in the art will readily recognize that additional probes and primers can be generated based upon the polynucleotide sequences provided by the disclosure. Referring to the primers and probes exemplified herein, a series of primers were designed using Primer Express Software (Applied Biosystems, Foster City, Calif.). The primers listed in SEQ ID NOs: 45-88 were designed, selected, and tested accordingly. In addition to the primers, reagents such as a dinucleotide triphosphate mixture having all four dinucleotide triphosphates (*e.g.*, dATP, dGTP, dCTP, and dTTP), a reverse transcriptase enzyme, and a thermostable DNA polymerase were used for RT-PCR. Additionally buffers, inhibitors and activators can

also be used for the RT-PCR process. Once the cDNA has been sufficiently amplified to a specified end point, the cDNA sample can be prepared for detection and quantitation. Though a number of detection schemes are contemplated, as will be discussed in more detail below, one method contemplated for detection of polynucleotides is fluorescence spectroscopy, and therefore labels suited to fluorescence spectroscopy are desirable for labeling polynucleotides. One example of such a fluorescent label is SYBR Green, though numerous related fluorescent molecules are known including, without limitation, DAPI, Cy3, Cy3.5, Cy5, CyS.5, Cy7, umbelliferone, fluorescein, fluorescein isothiocyanate (FITC), rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin.

[0065] In one embodiment of the disclosure, an oligonucleotide probe comprises a fragment of c-myc, CD44 antigen ("CD44"), cyclooxygenase 1 and 2("COX-1" and "COX-2"), cyclin D1, cyclin-dependent kinase inhibitor ("$p21^{cip/waf1}$"), interleukin 8 ("IL-8"), interleukin 8 receptor ("CXCR2"), osteopontin ("OPN"), melanoma growth stimulatory activity ("Groa/MGSA"), GRO3 oncogene ("Groγ"), macrophage colony stimulating factor 1 ("MCSF-1"), peroxisome proliferative activated receptor, alpha, delta and gamma ("PPAR-α, Δ and γ") and serum amyloid A1 ("SM 1") as set forth in Table 1.

[0066] Oligonucleotide probes and primers useful in the methods of the disclosure comprise at least 8 nucleotides of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, or 43 (including an oligonucleotide wherein T can be U) wherein the oligonucleotide specifically hybridizes to a polynucleotide sample from a subject comprising SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 , 25, 27, 29, 31, 33, 35, 37, 39, 41 and/or 43.

[0067] Any of the oligonucleotide primers and probes of the disclosure can be immobilized on a solid support. Solid supports are known to those skilled in the art and include the walls of wells of a reaction tray, test tubes, polystyrene beads, magnetic beads, nitrocellulose strips, membranes, microparticles such as latex particles, glass and the like. The solid support is not critical and can be selected by one skilled in the art. Thus, latex

particles, microparticles, magnetic or non-magnetic beads, membranes, plastic tubes, walls of microtiter wells, glass or silicon chips and the like are all suitable examples. Suitable methods for immobilizing oligonucleotides on a solid phase include ionic, hydrophobic, covalent interactions and the like. The solid support can be chosen for its intrinsic ability to attract and immobilize the capture reagent. The oligonucleotide probes or primers of the disclosure can be attached to or immobilized on a solid support individually or in groups of about 2-10,000 distinct oligonucleotides of the disclosure to a single solid support.

[0068] A substrate comprising a plurality of oligonucleotide primers or probes of the disclosure may be used either for detecting or amplifying targeted sequences. The oligonucleotide probes and primers of the disclosure can be attached in contiguous regions or at random locations on the solid support. Alternatively the oligonucleotides of the disclosure may be attached in an ordered array wherein each oligonucleotide is attached to a distinct region of the solid support which does not overlap with the attachment site of any other oligonucleotide. Typically, such oligonucleotide arrays are "addressable" such that distinct locations are recorded and can be accessed as part of an assay procedure. The knowledge of the location of oligonucleotides on an array make "addressable" arrays useful in hybridization assays. For example, the oligonucleotide probes can be used in an oligonucleotide chip such as those marketed by Affymetrix and described in U.S. Pat. No. 5,143,854; PCT publications WO 90/15070 and 92/10092, the disclosures of which are incorporated herein by reference. These arrays can be produced using mechanical synthesis methods or light directed synthesis methods which incorporate a combination of photolithographic methods and solid phase oligonucleotide synthesis.

[0069] The immobilization of arrays of oligonucleotides on solid supports has been rendered possible by the development of a technology generally referred to as "Very Large Scale Immobilized Polymer Synthesis" in which probes are immobilized in a high density array on a solid surface of a chip (see, *e.g.*, U.S. Patent Nos. 5,143,854; and 5,412,087 and in PCT Publications WO 90/15070,

WO 92/10092 and WO 95/11995, each of which are incorporated herein by reference), which describe methods for forming oligonucleotide arrays through techniques such as light-directed synthesis techniques.

[0070] In another aspect, an array of oligonucleotides complementary to subsequences of the target gene is used to determine the identity of the target, measure its amount, and detect differences between the target and a reference wild-type sequence.

[0071] Hybridization techniques can also be used to identify the biomarkers and/or polymorphisms of the disclosure and thereby determine or predict a colorectal cancer or gastrointestinal inflammatory disease or disorder. In this aspect, expression profiles or polymorphism(s) are identified based upon the higher thermal stability of a perfectly matched probe compared to the mismatched probe. The hybridization reactions may be carried out in a solid support (*e.g.*, membrane or chip) format, in which, for example, the target nucleic acids are immobilized on nitrocellulose or nylon membranes and probed with oligonucleotide probes of the disclosure. Any of the known hybridization formats may be used, including Southern blots, slot blots, "reverse" dot blots, solution hybridization, solid support based sandwich hybridization, bead-based, silicon chip-based and microtiter well-based hybridization formats.

[0072] Hybridization of an oligonucleotide probe to a target polynucleotide may be performed with both entities in solution, or such hybridization may be performed when either the oligonucleotide or the target polynucleotide is covalently or noncovalently affixed to a solid support. Attachment may be mediated, for example, by antibody-antigen interactions, poly-L-Lys, streptavidin or avidin-biotin, salt bridges, hydrophobic interactions, chemical linkages, UV cross-linking baking, etc. Oligonucleotides may be synthesized directly on the solid support or attached to the solid support subsequent to synthesis. Solid-supports suitable for use in detection methods of the disclosure include substrates made of silicon, glass, plastic, paper and the like, which may be formed, for example, into wells (as in 96-well plates), slides, sheets,

membranes, fibers, chips, dishes, and beads. The solid support may be treated, coated or derivatized to facilitate the immobilization of the allele-specific oligonucleotide or target nucleic acid.

[0073] In one aspect, a sandwich hybridization assay comprises separating the variant and/or wild-type target nucleic acid biomarker in a sample using a common capture oligonucleotide immobilized on a solid support and then contact with specific probes useful for detecting the variant and wild-type nucleic acids. The oligonucleotide probes are typically tagged with a detectable label.

[0074] Hybridization assays based on oligonucleotide arrays rely on the differences in hybridization stability of short oligonucleotides to perfectly matched and mismatched target variants. Efficient access to expression or polymorphic information is obtained through a basic structure comprising high-density arrays of oligonucleotide probes attached to a solid support (the chip) at selected positions. Each DNA chip can contain thousands to millions of individual synthetic DNA probes arranged in a grid-like pattern and miniaturized to the size of a dime or smaller. Such a chip may comprise oligonucleotides representative of both a wild-type and variant sequences.

[0075] Oligonucleotides of the disclosure can be designed to specifically hybridize to a target region of a polynucleotide. As used herein, specific hybridization means the oligonucleotide forms an anti-parallel double-stranded structure with the target region under certain hybridizing conditions, while failing to form such a structure when incubated with a different target polynucleotide or another region in the polynucleotide or with a polynucleotide lacking the desired locus under the same hybridizing conditions. Typically, the oligonucleotide specifically hybridizes to the target region under conventional high stringency conditions.

[0076] A nucleic acid molecule such as an oligonucleotide or polynucleotide is said to be a "perfect" or "complete" complement of another nucleic acid molecule if every nucleotide of one of the molecules is complementary to the nucleotide at the corresponding position of the other molecule. A nucleic acid molecule is "substantially complementary" to another molecule if it hybridizes

to that molecule with sufficient stability to remain in a duplex form under conventional low-stringency conditions. Conventional hybridization conditions are described, for example, in Sambrook *et al.*, Molecular Cloning, A Laboratory Manual, 2nd ed., Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (1989), and in Haymes *et al.*, Nucleic Acid Hybridization, A Practical Approach, IRL Press, Washington, D.C. (1985). While perfectly complementary oligonucleotides are used in most assays for detecting target polynucleotides or polymorphisms, departures from complete complementarity are contemplated where such departures do not prevent the molecule from specifically hybridizing to the target region. For example, an oligonucleotide primer may have a non-complementary fragment at its 5' or 3' end, with the remainder of the primer being complementary to the target region. Those of skill in the art are familiar with parameters that affect hybridization; such as temperature, probe or primer length and composition, buffer composition and salt concentration and can readily adjust these parameters to achieve specific hybridization of a nucleic acid to a target sequence.

[0077] A variety of hybridization conditions may be used in the disclosure, including high, moderate and low stringency conditions; see for example Maniatis *et al.*, Molecular Cloning: A Laboratory Manual, 2d Edition, 1989, and Short Protocols in Molecular Biology, ed. Ausubel, *et al.*, hereby incorporated by reference. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, stringent conditions are selected to be about 5-10°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target hybridize to the polyadenylated mRNA target sequence at equilibrium (as the target sequences are present

in excess, at T_m, 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30° C for short probes (*e.g.*, 10 to 50 nucleotides) and at least about 60° C for long probes (*e.g.*, greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of helix destabilizing agents such as formamide. The hybridization conditions may also vary when a non-ionic backbone, *i.e.*, PNA is used, as is known in the art. In addition, cross-linking agents may be added after target binding to cross-link, *i.e.*, covalently attach, the two strands of the hybridization complex.

[0078] A polymorphism in a target region of a gene may be assayed before or after amplification using one of several hybridization-based methods known in the art. Typically, allele-specific oligonucleotides are utilized in performing such methods. The allele-specific oligonucleotides may be used as differently labeled probe pairs, with one member of the pair showing a perfect match to one variant of a target sequence and the other member showing a perfect match to a different variant. In some embodiments, more than one polymorphism may be detected at once using a set of allele-specific oligonucleotides or oligonucleotide pairs. Typically, the members of the set have melting temperatures within 5 °C, and more typically within 2 °C, of each other when hybridizing to each of the polymorphic sites being detected.

[0079] In one aspect of for detection of polymorphisms, termed 4L tiled array, a set of four probes (A, C, G, T), typically 15-nucleotide oligomers in length is used. In each set of four probes, the perfect complement will hybridize more strongly than mismatched probes. Consequently, hybridization signals of the 15-mer probe set tiled array are perturbed by a single base change in the target sequence resulting in a characteristic loss of signal. Such techniques are particularly useful for detection of polymorphic regions in the biomarkers of the disclosure.

[0080] In another aspect, polymorphic regions of a biomarker of the disclosure may be identified. Diagnostic tests useful for

detecting polymorphic regions typically belong to two types: genotyping tests and haplotyping tests. A genotyping test simply provides the status of a variance or variances in a subject. For example, suppose nucleotide 150 of hypothetical gene X on an autosomal chromosome is an adenine (A) or a guanine (G) base. The possible genotypes in an individual with the gene are AA, AG or GG at nucleotide 150 of gene X.

[0081] In a haplotyping test there is at least one additional variance in gene X, say at nucleotide 810, which varies in the population as cytosine (C) or thymine (T). Thus a particular copy of gene X may have any of the following combinations of nucleotides at positions 150 and 810: 150A-810C, 150A-810T, 150G-810C or 150G-810T. Each of the four possibilities is a unique haplotype. If the two nucleotides interact in either RNA or protein, then knowing the haplotype can be important. The point of a haplotyping test is to determine the haplotypes present in a DNA or cDNA sample (*e.g.* from a subject).

[0082] Methods and compositions of the disclosure are useful for diagnosing or determining the risk of developing a colorectal cancer or gastrointestinal inflammatory disease or disorder. Such tests can be performed using DNA or RNA samples collected from blood, cells, tissue scrapings or other cellular materials, and can be performed by a variety of methods including, but not limited to, hybridization with biomarker-specific probes, enzymatic mutation detection, chemical cleavage of mismatches, mass spectrometry, PCR or DNA sequencing, including minisequencing. Diagnostic tests may involve a panel of from one or more genes, genetic markers (gene expression profiles), often on a solid support, or using PCR techniques, which enables the simultaneous determination of more than one variance in one or more genes or expression of one or more genes.

[0083] A target biomarker or region(s) thereof (*e.g.*, containing a polymorphism of interest) may be amplified using any oligonucleotide-directed amplification method including, but not limited to, polymerase chain reaction (PCR) (U.S. Pat. No. 4,965,188), ligase chain reaction (LCR) (Barany *et al.*, Proc. Natl. Acad. Sci. USA 88:189-93 (1991); WO 90/01069), and oligonucleotide

ligation assay (OLA) (Landegren *et al.*, Science 241:1077-80 (1988)). Other known nucleic acid amplification procedures may be used to amplify the target region(s) including transcription-based amplification systems (U.S. Pat. No. 5,130,238; European Patent No. EP 329,822; U.S. Pat. No. 5,169,766; WO 89/06700) and isothermal methods (Walker *et al.*, Proc. Natl. Acad. Sci. USA 89:392-6 (1992)).

[0084] Ligase Chain Reaction (LCR) techniques can be used and are particularly useful for detection of polymorphic variants. LCR occurs only when the oligonucleotides are correctly base-paired. The Ligase Chain Reaction (LCR), which utilizes the thermostable Taq ligase for ligation amplification, is useful for interrogating loci of a gene (*e.g.*, comprising SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41 or 43). A method of DNA amplification similar to PCR, LCR differs from PCR because it amplifies the probe molecule rather than producing amplicon through polymerization of nucleotides. Two probes are used per each DNA strand and are ligated together to form a single probe. LCR uses both a DNA polymerase enzyme and a DNA ligase enzyme to drive the reaction. Like PCR, LCR requires a thermal cycler to drive the reaction and each cycle results in a doubling of the target nucleic acid molecule. LCR can have greater specificity than PCR. The elevated reaction temperatures permit the ligation reaction to be conducted with high stringency. Where a mismatch occurs, ligation cannot be accomplished. For example, a primer based upon a target gene or gene variant is synthesized in two fragments and annealed to the template with possible mutation at the boundary of the two primer fragments (*i.e.*, the underlined nucleotide above would be found at the 5' or 3' end of the oligonucleotide). A ligase ligates the two primers if they match exactly to the template sequence.

[0085] In one embodiment, the two hybridization probes are designed each with a target specific portion. The first hybridization probe is designed to be substantially complementary to a first target domain of a target polynucleotide (*e.g.*, a polynucleotide fragment) and the second hybridization probe is substantially complementary to a second target domain of a target

polynucleotide (*e.g.*, a polynucleotide fragment). In general, each target specific sequence of a hybridization probe is at least about 5 nucleotides long, with sequences of about 15 to 30 being typical and 20 being especially common. In one embodiment, the first and second target domains are directly adjacent, *e.g.*, they have no intervening nucleotides. In this embodiment, at least a first hybridization probe is hybridized to the first target domain and a second hybridization probe is hybridized to the second target domain. If perfect complementarity exists at the junction, a ligation structure is formed such that the two probes can be ligated together to form a ligated probe. If this complementarity does not exist (due to mismatch based upon a variant), no ligation structure is formed and the probes are not ligated together to an appreciable degree. This may be done using heat cycling, to allow the ligated probe to be denatured off the target polynucleotide such that it may serve as a template for further reactions. The method may also be done using three hybridization probes or hybridization probes that are separated by one or more nucleotides, if dNTPs and a polymerase are added (this is sometimes referred to as "Genetic Bit" analysis).

[0086] Analysis of point mutations (*e.g.*, polymorphic variants) in DNA can also be carried out by using the polymerase chain reaction (PCR) and variations thereof. Mismatches can be detected by competitive oligonucleotide priming under hybridization conditions where binding of the perfectly matched primer is favored. In the amplification refractory mutation system technique (ARMS), primers are designed to have perfect matches or mismatches with target sequences either internal or at the 3' residue (Newton *et al.*, Nucl. Acids. Res. 17:2503-2516 (1989)). Under appropriate conditions, only the perfectly annealed oligonucleotide functions as a primer for the PCR reaction, thus providing a method of discrimination between normal and variant sequences.

[0087] Single nucleotide primer-guided extension assays can also be used, where the specific incorporation of the correct base is provided by the fidelity of a DNA polymerase. Detecting the nucleotide or nucleotide pair at a polymorphic site of interest may also be determined using a mismatch detection technique including,

but not limited to, the RNase protection method using riboprobes (Winter *et al.*, Proc. Natl. Acad. Sci. USA 82:7575 (1985); Meyers *et al.*, Science 230:1242 (1985)) and proteins which recognize nucleotide mismatches, such as the *E. coli* mutS protein (Modrich, Ann. Rev. Genet. 25:229-53 (1991)). Alternatively, variant alleles can be identified by single strand conformation polymorphism (SSCP) analysis (Orita *et al.*, Genomics 5:874-9 (1989); Humphries *et al.*, in MOLECULAR DIAGNOSIS OF GENETIC DISEASES, Elles, ed., pp. 321-340, 1996) or denaturing gradient gel electrophoresis (DGGE) (Wartell *et al.*, Nucl. Acids Res. 18:2699-706 (1990); Sheffield *et al.*, Proc. Natl. Acad. Sci. USA 86:232-6 (1989)).

[0088] A polymerase-mediated primer extension method may also be used to identify the polymorphism(s). Several such methods have been described in the patent and scientific literature and include the "Genetic Bit Analysis" method (WO 92/15712) and the ligase/polymerase mediated genetic bit analysis (U.S. Pat. No. 5,679,524. Related methods are disclosed in WO 91/02087, WO 90/09455, WO 95/17676, and U.S. Pat. Nos. 5,302,509 and 5,945,283. Extended primers containing the complement of the polymorphism may be detected by mass spectrometry as described in U.S. Pat. No. 5,605,798. Another primer extension method is allele-specific PCR (Ruano *et al.*, 1989, supra; Ruano *et al.*, 1991, supra; WO 93/22456; Turki *et al.*, J. Clin. Invest. 95:1635-41 (1995)).

[0089] Another technique, which may be used to analyze gene expression and polymorphisms, includes multicomponent integrated systems, which miniaturize and compartmentalize processes such as PCR and capillary electrophoresis reactions in a single functional device. An example of such technique is disclosed in U.S. Pat. No. 5,589,136, the disclosure of which is incorporated herein by reference in its entirety, which describes the integration of PCR amplification and capillary electrophoresis in chips.

[0090] Quantitative PCR and digital PCR can be used to measure the level of a polynucleotide in a sample. Digital Polymerase Chain Reaction (digital PCR, dPCR or dePCR) can be used to directly quantify and clonally amplify nucleic acids including DNA, cDNA or RNA. Digital PCR amplifies nucleic acids by temperature cycling of a nucleic acid molecule with a DNA polymerase. The reaction is

typically carried out in the dispersed phase of an emulsion capturing each individual nucleic acid molecule present in a sample within many separate chambers or regions prior to PCR amplification. A count of chambers containing detectable levels of PCR end-product is a direct measure of the absolute nucleic acids quantity.

[0091] Quantitative polymerase chain reaction (qPCR) is a modification of the polymerase chain reaction and real-time quantitative PCR are useful for measuring the amount of DNA after each cycle of PCR by use of fluorescent markers or other detectable labels. Quantitative PCR methods use the addition of a competitor RNA (for reverse-transcriptase PCR) or DNA in serial dilutions or co-amplification of an internal control to ensure that the amplification is stopped while in the exponential growth phase.

[0092] Modifications of PCR and PCR techniques are routine in the art and there are commercially available kits useful for PCR amplification.

[0093] The detectable label may be a radioactive label or may be a luminescent, fluorescent of enzyme label. Indirect detection processes typically comprise probes covalently labeled with a hapten or ligand such as digoxigenin (DIG) or biotin. In one aspect, following the hybridization step, the target-probe duplex is detected by an antibody- or streptavidin-enzyme complex. Enzymes commonly used in DNA diagnostics are horseradish peroxidase and alkaline phosphatase. Direct detection methods include the use of fluorophor-labeled oligonucleotides, lanthanide chelate-labeled oligonucleotides or oligonucleotide-enzyme conjugates. Examples of fluorophor labels are fluorescein, rhodamine and phthalocyanine dyes.

[0094] Examples of detection modes contemplated for the disclosed methods include, but are not limited to, spectroscopic techniques, such as fluorescence and UV-Vis spectroscopy, scintillation counting, and mass spectroscopy. Complementary to these modes of detection, examples of labels for the purpose of detection and quantitation used in these methods include, but are not limited to, chromophoric labels, scintillation labels, and mass labels. The expression levels of polynucleotides and polypeptides

measured using these methods may be normalized to a control established for the purpose of the targeted determination.

[0095] Label detection will be based upon the type of label used in the particular assay. Such detection methods are known in the art. For example, radioisotope detection can be performed by autoradiography, scintillation counting or phosphor imaging. For hapten or biotin labels, detection is with an antibody or streptavidin bound to a reporter enzyme such as horseradish peroxidase or alkaline phosphatase, which is then detected by enzymatic means. For fluorophor or lanthanide-chelate labels, fluorescent signals may be measured with spectrofluorimeters with or without time-resolved mode or using automated microtitre plate readers. With enzyme labels, detection is by color or dye deposition (p-nitropheny phosphate or 5-bromo-4-chloro-3-indolyl phosphate/nitroblue tetrazolium for alkaline phosphatase and 3,3'-diaminobenzidine-$NiCl_2$ for horseradish peroxidase), fluorescence (*e.g.*, 4-methyl umbelliferyl phosphate for alkaline phosphatase) or chemiluminescence (the alkaline phosphatase dioxetane substrates LumiPhos 530 from Lumigen Inc., Detroit Mich. or AMPPD and CSPD from Tropix, Inc.). Chemiluminescent detection may be carried out with X-ray or polaroid film or by using single photon counting luminometers.

[0096] In another aspect of this disclosure, expression levels of proteins comprising SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42 and/or 44 can be measured and quantitated using techniques known in the art including, for example, Western blots, ELISA assays and the like. The term "polypeptide" or "polypeptides" is used interchangeably with the term "protein" or "proteins" herein.

[0097] In another embodiment, a method for protein expression profiling comprises using one or more (*e.g.*, a plurality of) antibodies to one or more biomarkers for measuring targeted polypeptide levels from a biological sample. In one embodiment contemplated for the method, the antibodies for the panel are bound to a solid support. The method for protein expression profiling may use a second antibody having specificity to some portion of the bound polypeptide. Such a second antibody may be detectably

labeled with molecules useful for detection and quantitation of the bound polypeptides. Additionally, other reagents are contemplated for detection and quantitation including, for example, small molecules such as cofactors, substrates, complexing agents, and the like, or large molecules, such as lectins, peptides, olionucleotides, and the like. Such moieties may be either naturally occurring or synthetic.

[0098] The disclosure further contemplates, antibodies capable of specifically binding to a biomarker polypeptides encoded in proper frame, based upon transcriptional and translational starts, of the above-identified polynucleotide biomarker sequences (*e.g.*, comprising SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, or 43). The disclosure thus includes isolated, purified, and recombinant polypeptides comprising a contiguous span of at least 4 amino acids, typically at least 6, more commonly at least 8 to 10 amino acids encoded by a polynucleotide comprising SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 37, 39, 41, or 43.

[0099] The disclosure also contemplates the use of immunoassay techniques for measurement of polypeptide biomarkers identified herein. The polypeptide biomarker can be isolated and used to prepare antisera and monoclonal antibodies that specifically detect a biomarker gene product. Mutated gene products also can be used to immunize animals for the production of polyclonal antibodies. Recombinantly produced peptides can also be used to generate antibodies. For example, a recombinantly produced fragment of a polypeptide can be injected into a mouse along with an adjuvant so as to generate an immune response. Murine immunoglobulins which bind the recombinant fragment with a binding affinity of at least $1x10^{7}$ M^{-1} can be harvested from the immunized mouse as an antiserum, and may be further purified by affinity chromatography or other means. Additionally, spleen cells are harvested from the mouse and fused to myeloma cells to produce a bank of antibody-secreting hybridoma cells. The bank of hybridomas can be screened for clones that secrete immunoglobulins which bind the recombinantly produced fragment with an affinity of at least $1x10^{6}$ M^{-1}. More specifically, immunoglobulins that selectively bind to the variant polypeptides

but poorly or not at all to wild-type polypeptides are selected, either by pre-absorption with wild-type proteins or by screening of hybridoma cell lines for specific idiotypes that bind the variant, but not wild-type, polypeptides.

[00100] Polynucleotides capable of expressing the polypeptides can be generated using techniques skilled in the art based upon the identified sequences herein. Such polynucleotides can be expressed in hosts, wherein the polynucleotide is operably linked to (*i.e.*, positioned to ensure the functioning of) an expression control sequence. Expression vectors are typically replicable in the host organisms either as episomes or as an integral part of the host chromosome. Expression vectors can contain selection markers (*e.g.*, markers based on tetracyclin resistance or hygromycin resistance) to permit detection and/or selection of those cells transformed with the desired polynucleotide.

[00101] Polynucleotides encoding a variant polypeptide may include sequences that facilitate transcription and translation of the coding sequences such that the encoded polypeptide product is produced. Construction of such polynucleotides is known in the art. For example, such polynucleotides can include a promoter, a transcription termination site (polyadenylation site in eukaryotic expression hosts), a ribosome binding site, and, optionally, an enhancer for use in eukaryotic expression hosts, and, optionally, sequences necessary for replication of a vector.

[00102] Prokaryotes can be used as host cells for the expression of a variant polypeptides, such techniques are known in the art. Other microbes, such as yeast, may also be used for expression. In addition to microorganisms, mammalian tissue cell culture may also be used to express and produce polypeptides of the disclosure. Eukaryotic cells useful in the methods of the disclosure include the CHO cell lines, various COS cell lines, HeLa cells, myeloma cell lines, Jurkat cells, and so forth. Expression vectors for these cells can include expression control sequences, such as an origin of replication, a promoter, an enhancer, an necessary information processing sites, such as ribosome binding sites, RNA splice sites, polyadenylation sites, and transcriptional terminator sequences.

[00103] The techniques for polynucleotide cloning and expression are useful in the disclosure for the generation of probes capable of hybridizing to polynucleotide biomarkers or the generation of antibodies useful for binding polypeptide biomarkers of the disclosure.

[00104] In further methods, peptides, drugs, fatty acids, lipoproteins, or small molecules which interact with a biomarker (*e.g.*, a polynucleotide or polypeptide, protein, or a fragment comprising a contiguous span of at least 4 amino acids, at least 6 amino acids, or typically at least 8 to 10 amino acids or more of sequences corresponding to the biomarkers herein) can be used as detection agents for measuring biomarkers. The molecule to be tested for binding is labeled with a detectable label, such as a fluorescent, radioactive, or enzymatic tag. After removal of non-specifically bound molecules, bound molecules are detected using appropriate means.

[00105] These results, with reference to the figures and specific examples below, demonstrate that it is possible to sample cells through a minimally invasive swabbing collection method from an area distant from a cancerous lesion, but capable of indicating a non-normal colon condition. In that regard, samples taken either minimally invasively or non-invasively would render samples that could be analyzed using the disclosed panel of biomarkers. Such non-invasive procedures not only reduce the cost of determination of CRC, but reduce the discomfort and risk associated with current methodology. All these factors together increase the attractiveness of regular testing, and hence patient compliance. Increased patient compliance, coupled with an effective determination for CRC, enhance the prospects for early detection, and enhanced survival rates.

[00106] Table 15 below demonstrates the differences in expression profiles based upon biomarkers of the disclosure. FHSH refers to family and self history of the subject. FHSH subjects lacked a history of polyps. In addition, FHSH subject can lack a history of gastrointestinal diseases or disorders. As referenced in table 15, "Others" refer to subjects that have a history of gastrointestinal diseases or disorders. Accordingly, in one aspect

of the disclosure, a predictive biomarker for gastrointestinal inflammatory disease or disorder would include detecting a change in expression of IL-8, CD44, c-myc, and/or P21, which all show larger changes (*e.g.*, about 19, 63, 50 and 56%, respectively, relative to controls). It is important to note that a change in expression of a biomarker of the disclosure need not necessarily be an increase in expression relative to a control. Rather, a change can be an increase or decrease relative to a control so long as the change represents a statistically significant difference relative to the control. In one aspect, the change is at least 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75% or more in an increase or decrease relative to a control. Where a panel of biomarkers are used in the detection of a disease or disorder, a smaller change relative to a control can be indicative of the disease or disorder or risk thereof in comparison to a change in each biomarker alone. A statistician of skill in the art will be capable of identifying statistically significant differences in a biomarker or panel of biomarkers relative to a control value(s).

[00107] In principle, the larger the number of genes used, the more sensitive the analysis is will be. The panel can comprise from 3 to fifteen or sixteen genes or biomarkers. In one aspect, the panel comprises 15 ro 16 genes or biomarkers. However, for individuals with polyps or with history of cancer, the specificity is somewhat less, and fine-tuning the analysis by adding to or otherwise modifying the gene panel increases specificity. As discussed below, the procedure involves determining which genes in the panel make the largest contribution to significance.

[00108] Using the methods described herein, research on APCmin mice, identified a panel of mRNAs with highly up-regulated activities associated with colorectal cancer. Similar genes were seen upregulated in human samples. While the pathologist would describe the staging of the cancer in terms of depth of invasion and the presence or absence of lymph node involvement, among other variable, with the usual comment that the margins were clear, gene expression data demonstrated that the margins showed highly up-regulated mRNAs, and these values were high all over the entire specimen, not just adjacent to the cancer itself. Case after case

of such resected colon cancer specimens showed the identical data. The panel of 16 selected mRNAs comprised of many different metabolic pathways resulted in a new panel useful for diagnostics.

[00109] These same mRNAs showed minimal activity in colons with no polyp or cancer. These patients were males and females, Caucasians and Asians and the results were the same, very low values with normal colons.

[00110] In patients with colon cancers and in many patients with pre-malignant polyps, these values were high not only in the region of the cancer or the polyp, but also far away from these lesions, as far away as the rectum. The rectal biopsy values were as abnormal even when the lesion was in the ascending colon or cecum.

[00111] Ninety patients were examined to demonstrate the methods and compositions of the disclosure. Although the activities of the panel of 16 genes may vary slightly between the two samples, they essentially yield the same results. This is probably due to the slight difference in the cells so collected, with the biopsy samples being deeper into the rectal mucosa and the smear samples coming entirely from the surface of the rectal lining. Thus, a simple rectal smear through an ordinary anoscope, without bowel preparation, will give a glimpse of what the rest of the colon looks like. Cancer cases had extremely high values. The data strongly support that a highly up-regulated mRNA activity in a selected panel of the disclosure from a simple rectal smear correlates with a colorectal cancer anywhere in the colon.

[00112] In one of the study population 52% were males and 48% females. 43% were Caucasians, 52% were Asians and 4% were Africian American. For the patients with a positive family history of colorectal cancer, some showed elevated activities and some did not. For the patients with polyps, some showed elevated activities, particularly those with significant polyps 2 cm or larger or with villous component. Most of the patients with hyperplastic polyps showed normal activities, although a few had abnormal values. Interestingly enough, for patients who simply had intermittently rectal bleeding without any risk factors, some showed abnormal levels and some did not. Those patients with no polyp or cancer and with no risk factors had very low values. Lastly, there were three

patients with very high values without a polyp or a cancer. One had Crohns' disease involving the sigmoid colon and two had Barrett's esophagus.

[00113] In another aspect, the disclosure provides methods of early detection or diagnosis of a colorectal cancer or gastrointestinal inflammatory disease or disorder based upon measurement of any of the biomarkers in tables 3-14 by rectal, colon, or buccal swabs. This method can be followed by a determination at a later time by measuring the same, one or more additional genes, or one or more additional biomarker panels. For example, early detection or diagnosis can be based upon screening changes in any one or more of the biomarkers described, wherein a change in a biomarker's expression (*e.g.*, IL-8, P21, c-myc, and/or CD44) realtive to a control is indicative of a gastrointestinal inflammatory disease or disorder or the risk of acquiring an gastrointestinal inflammatory disease or disorder; following initial diagnosis or prediction the same or different makers (*e.g.*, IL-8) can be measured to determine the prognosis or development of a disease. The data below indicate, for example, that the biomarker IL-8 and OPN may be indicative of later stage development of a gastrointestinal disease or disorder.

Table 15

	Swabs **FHSH,n=16** **p<0.0000**	***Swabs*** **Others, n=9** **p<0.0000**	***Biopsies*** **FHSH, n=17** **p<0.0001**	***Biopsies*** **Others, n=8** **p<0.0001**
Overall				
CXCR2	19%	56%	57%	38%
OPN	38	44	18	63
COX1	42	33	18	13
PPARα	15	22	12	13
COX2	38	44	12	13
Groα	50	56	29	25
Groγ	42	56	17	25
IL8	23	67	12	13
PPARγ	31	33	17	25
P21	31	78	12	25
cMyc	38	56	29	13
CD44	46	67	17	13
mCSF-1	35	33	0	0
cycD	31	44	12	0
PPARδ	31	56	24	50
SAA1	27	22	12	25

[00114] In other embodiments, the computer-readable medium for determine a risk, prognosis or diagnosis of a gastrointestinal disorder or disease (e.g., an IBD, polyp or cancer) comprises instructions to apply a statistical process to a data set comprising a biomarker profile optionally in combination with a symptom profile provided by a technician, nurse or physician, which indicates the presence or severity of at least one symptom in the individual to produce a statistically derived decision classifying the sample as a (i) non-colorectal cancer gastrointestinal disease or disorder; (ii) a polyp stage disease or disorder or (iii) a colorectal cancer stage disease or disorder, based upon the biomarker profile or the biomarker profile and the symptom profile.

[00115] In another embodiment, a computer-readable medium including code for controlling one or more processors to classify whether a sample from an individual is associated (i) non-colorectal cancer gastrointestinal disease or disorder; (ii) a polyp stage disease or disorder or (iii) a colorectal cancer stage disease or disorder comprising: (a) instructions to apply a first statistical process to a data set comprising a biomarker profile to produce a statistically derived decision classifying the sample as (i) non-colorectal cancer gastrointestinal disease or disorder; (ii) a polyp stage disease or disorder or (iii) a colorectal cancer stage disease or disorder based upon the biomarker profile; and if the sample is classified as a (i) non-colorectal cancer gastrointestinal disease or disorder; (ii) a polyp stage disease or disorder or (iii) a colorectal cancer stage disease or disorder, (b) instructions to apply a second statistical process to the same or different data set to produce a second statistically derived decision classifying the (i) non-colorectal cancer gastrointestinal disease or disorder; (ii) a polyp stage disease or disorder or (iii) a colorectal cancer stage disease or disorder.

[00116] In another embodiment, a process can use a computer to apply a second statistic approach to a biomarker panel measurement based upon a earlier determine criteria (e.g., if a polyp diagnosis, then apply colorectal biomarker panel measurements and statistics; if a FHSH disposition then apply polyp biomarker panel measurements and statistics).

[00117] In yet another embodiment, the methods and systems of the disclosure provide for classifying whether a sample from an individual is associated with (i) non-colorectal cancer gastrointestinal disease or disorder; (ii) a polyp stage disease or disorder or (iii) a colorectal cancer stage disease or disorder, the system comprising: (a) a data acquisition module configured to produce a data set comprising a biomarker profile, wherein the biomarker profile indicates the presence or level of at least one biomarker in the sample; (b) a data processing module configured to process the data set by applying a statistical process to the data set to produce a statistically derived decision classifying the sample as an (i) non-colorectal cancer gastrointestinal disease or disorder; (ii) a polyp stage disease or disorder or (iii) a colorectal cancer stage disease or disorder sample based upon the diagnostic marker profile; and (c) a display module configured to display the statistically derived decision.

[00118] In certain instances, the statistical algorithm is a learning statistical classifier system. The learning statistical classifier system can be selected from the group consisting of a random forest (RF), classification and regression tree (C&RT), boosted tree, neural network (NN), support vector machine (SVM), general chi-squared automatic interaction detector model, interactive tree, multiadaptive regression spline, machine learning classifier, and combinations thereof. Preferably, the learning statistical classifier system is a tree-based statistical algorithm (e.g., RF, C&RT, etc.) and/or a NN (e.g., artificial NN, etc.).

[00119] In certain instances, the statistical algorithm is a single learning statistical classifier system. Typically, the single learning statistical classifier system comprises a tree-based statistical algorithm such as a RF or C&RT. As a non-limiting example, a single learning statistical classifier system can be used to classify the sample as an (i) non-colorectal cancer gastrointestinal disease or disorder; (ii) a polyp stage disease or disorder or (iii) a colorectal cancer stage disease or disorder based upon a prediction or probability value and the presence or level of at least one biomarker (or panel of biomarkers), alone or in combination with the presence or severity of at least one

symptom (i.e., symptom profile). The use of a single learning statistical classifier system typically classifies the sample as an (i) non-colorectal cancer gastrointestinal disease or disorder; (ii) a polyp stage disease or disorder or (iii) a colorectal cancer stage disease or disorder with a sensitivity, specificity, positive predictive value, negative predictive value, and/or overall accuracy of at least about 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%.

[00120] In some instances, the data obtained from using the learning statistical classifier system or systems can be processed using a processing algorithm. Such a processing algorithm can be selected, for example, from the group consisting of a multilayer perceptron, backpropagation network, and Levenberg-Marquardt algorithm. In other instances, a combination of such processing algorithms can be used, such as in a parallel or serial fashion.

[00121] In a further embodiment, the methods of the disclosure further comprise sending the (i) non-colorectal cancer gastrointestinal disease or disorder; (ii) a polyp stage disease or disorder or (iii) a colorectal cancer stage disease or disorder classification results to a clinician, e.g., a gastroenterologist or a general practitioner. In another embodiment, the methods provides a diagnosis or prognosis in the form of a probability that the individual has (i) non-colorectal cancer gastrointestinal disease or disorder; (ii) a polyp stage disease or disorder or (iii) a colorectal cancer stage disease or disorder. For example, the individual can have about a 0%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or greater probability of having (i) non-colorectal cancer gastrointestinal disease or disorder; (ii) a polyp stage disease or disorder or (iii) a colorectal cancer stage disease or disorder.

[00122] In another embodiment, a method of the disclosure provides a method for classifying whether a sample from an individual is associated with (i) a polyp stage disease or disorder comprising:(a) determining a biomarker profile by detecting the presence or level of at least one biomarker in the sample associated with polyps; (b) classifying the sample as a polyp

sample using a first statistical algorithm based upon the biomarker profile; and if the sample is classified as a polyp sample, (c) classifying the polyp sample as an polyp or colorectal cancer stage sample using a second statistical algorithm based upon a biomarker profile by detecting the presence or level of at least one biomarker in the sample associated with colorectal cancer (e.g., by swab or bioposy) and classifying the sample as a colorectal cancer sample suing a second statistical algorithm based upon a colorectal cancer biomarker panel.

[00123] One skilled in the art will appreciate that the presence or level of a plurality of biomarkers can be determined simultaneously or sequentially, using, for example, an aliquot or dilution of the individual's sample. As described above, the level of a particular biomarker in the individual's sample is generally considered to be elevated when it is at least about 25%, 50%, 75%, 100%, 125%, 150%, 175%, 200%, 250%, 300%, 350%, 400%, 450%, 500%, 600%, 700%, 800%, 900%, or 1000% greater than the level of the same marker in a comparative sample or population of samples (e.g., greater than a median level). Similarly, the level of a particular diagnostic marker in the individual's sample is typically considered to be lowered when it is at least about 5%,10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95% less than the level of the same marker in a comparative sample or population of samples (e.g., less than a median level).

[00124] Methods and kits for the polynucleotide and polypeptide expression profiling for the panel of molecular markers are also contemplated as part of the present disclosure.

[00125] In one embodiment, a kit for gene expression profiling comprises the reagents and instructions necessary for the gene expression profiling of the biomarkers or biomarker panel. Thus, for example, the reagents may include primers, enzymes, and other reagents for the preparation, detection, and quantitation of cDNAs for the claimed panel of biomarkers. The primers listed in SEQ ID NOs: 45-88 are particularly suited for use in gene expression profiling using RT-PCR based on the claimed panel. The primers listed in SEQ ID NOs: 45-88 were specifically designed, selected,

and tested accordingly. In addition to the primers, reagents such as dinucleotide triphosphate comprising dinucleotide triphosphates (*e.g.*, dATP, dGTP, dCTP, and dTTP), reverse transcriptase, and a thermostable DNA polymerase. Additionally buffers, inhibitors and activators used for the RT-PCR process are suitable reagents for inclusion in the kit embodiment. Once the cDNA has been sufficiently amplified to a specified end point, the cDNA sample must be prepared for detection and quantitation. One method contemplated for detection of polynucleotides is fluorescence spectroscopy using fluorescent moieties or labels that are suited to fluorescence spectroscopy are desirable for labeling polynucleotides and may also be included in reagents of the kit embodiment.

[00126] In one embodiment, the disclosure provides a kit useful for identifying biomarkers indicative of a gastrointestinal disease or disorder. For example, the kit of the disclosure can comprise one or more oligonucleotides designed for identifying alleles and/or biomarkers of the disclosure. In another embodiment, the kit further comprises a manual with instructions for (a) performing one or more reactions on a human nucleic acid sample to identify biomarkers and/or alleles present in the subject.

[00127] The oligonucleotides in a kit of the disclosure may also be immobilized on or synthesized on a solid surface such as a microchip, bead, or glass slide (see, *e.g.*, WO 98/20020 and WO 98/20019). Such immobilized oligonucleotides may be used in a variety of detection assays, including but not limited to, probe hybridization and polymerase extension assays. Immobilized oligonucleotides useful in practicing the disclosure may comprise an ordered array of oligonucleotides designed to rapidly screen a nucleic acid sample.

[00128] Kits of the disclosure may also contain other components such as hybridization buffer (*e.g.*, where the oligonucleotide probes) or dideoxynucleotide triphosphates (ddNTPs; *e.g.*, for primer extension). In one embodiment, the set of oligonucleotides consists of primer-extension oligonucleotides. The kit may also contain a polymerase and a reaction buffer optimized for primer-extension mediated by the polymerase. Kits may also include

detection reagents, such as biotin- or fluorescent-tagged oligonucleotides or ddNTPs and/or an enzyme-labeled antibody and one or more substrates that generate a detectable signal when acted on by the enzyme. It is also contemplated that the above described methods and compositions of the disclosure may be utilized in combination with other biomarker techniques.

[00129] Nucleic acid samples, for example for use in variance identification, can be obtained from a variety of sources as known to those skilled in the art, or can be obtained from genomic or cDNA sources by known methods.

[00130] In another embodiment, a kit for protein expression profiling comprises the reagents and instructions necessary for protein expression profiling of a polypeptide biomarker panel. Thus, in this embodiment, the kit for protein expression profiling includes supplying an antibody panel based on a panel of biomarkers for measuring targeted polypeptide levels from a biological sample. One embodiment contemplated for such a panel includes the antibody panel bound to a solid support. Additionally, the reagents included with the kit for protein expression profiling may use a second antibody having specificity to some portion of the bound polypeptide. Such a second antibody may be labeled with molecules useful for detection and quantitation of the bound polypeptides.

[00131] Generally, the diagnostic test of the disclosure involves determining whether an individual has a variance or variant form of a gene or a change in expression.

[00132] Integrated systems can be envisaged mainly when microfluidic systems are used. These systems comprise a pattern of microchannels designed onto a glass, silicon, quartz, or plastic wafer included on a microchip. The movements of the samples are controlled by electric, electroosmotic or hydrostatic forces applied across different areas of the microchip. The microfluidic system may integrate nucleic acid amplification, microsequencing, capillary electrophoresis and a detection method such as laser-induced fluorescence detection.

[00133] It is also contemplated that the gene expression profile may be transmitted to a remote location for analysis. For example, changes in a detectable signal related to gene expression from a

first time and a second time are communicated to a remote location for analysis.

[00134] The digital representation of the detectable signal is transmittable over any number of media. For example, such digital data can be transmitted over the Internet in encrypted or in publicly available form. The data can be transmitted over phone lines, fiber optic cables or various air-wave frequencies. The data are then analyzed by a central processing unit at a remote site, and/or archived for compilation of a data set that could be mined to determine, for example, changes with respect to historical mean "normal" values of a genetic expression profile of a subject.

[00135] Embodiments of the disclosure include systems (*e.g.*, internet based systems) particularly computer systems which store and manipulate the data corresponding to the detectable signal obtained an expression profile. As used herein, "a computer system" refers to the hardware components, software components, and data storage components used to analyze the digital representative of an expression profile or plurality of profiles. The computer system typically includes a processor for processing, accessing and manipulating the data. The processor can be any well-known type of central processing unit.

[00136] Typically the computer system is a general purpose system that comprises the processor and one or more internal data storage components for storing data, and one or more data retrieving devices for retrieving the data stored on the data storage components. A skilled artisan can readily appreciate that any one of the currently available computer systems are suitable.

[00137] In one particular embodiment, the computer system includes a processor connected to a bus which is connected to a main memory (preferably implemented as RAM) and one or more internal data storage devices, such as a hard drive and/or other computer readable media having data recorded thereon. In some embodiments, the computer system further includes one or more data retrieving device for reading the data stored on the internal data storage devices.

[00138] The data retrieving device may represent, for example, a floppy disk drive, a compact disk drive, a magnetic tape drive, or

a modem capable of connection to a remote data storage system (*e.g.*, via the internet) and the like. In some embodiments, the internal data storage device is a removable computer readable medium such as a floppy disk, a compact disk, a magnetic tape, and the like, containing control logic and/or data recorded thereon. The computer system may advantageously include or be programmed by appropriate software for reading the control logic and/or the data from the data storage component once inserted in the data retrieving device.

EXAMPLES

[00139] The genes in the expression panel fall into four major groups: 1) APC/b-catenin pathway, including c-myc, cyclin D1, and proliferating peroxisome activating receptor (PPAR alpha, delta and gamma); 2) NF-kB/inflammation pathway, including the growth-related oncogenes (Gro)-alpha and gamma osteopontin (OPN), and colony-stimulating factor (M-CSF-1), cyclo-oxygenases (COX)-1 and 2, interleukin-8 (IL-8), and the cytokine receptor CXCR2; 3) cell cycle/transcription factors, including p21, cyclin D1, c-myc, PPAR alpha, delta and gamma and 4) cell communication signals, including IL-8, PPAR alpha, delta and gamma, CXCR2, CD44, and OPN. Most of these genes are shown to be up-regulated in human colon cancers, though a few, such as the p21, as well as PPAR alpha, delta and gamma are down-regulated.

[00140] The disclosure also provides information comparing rectal swabs vs. biopsies as a means of tissue collection, in about 90 individuals, 37 individuals with history, 25 individuals with polyps (with or without history), and 23 controls with no polyps, no family or self history of cancer, and no known obvious upper GI problems. In this 90 patient study there was no cancer in situ case, 5 individuals scheduled for surgery due to colon cancer were swabbed.

[00141] The methods compare gene expression values of normal appearing mucosa of individuals or a group with cancer or cancer risk with values from controls. The statistical approach generally begins with a global multivariate analysis of variance (ANOVA), that takes into account correlations among the expression levels of different genes. This type of analysis controls the false positive

rate by providing a single test of whether the expression patterns, based on all the genes in the subset, differ between groups or individuals. If the global test is significant for a particular individual or for a particular group, a univariate test was then used to determine which genes are contributing to the global difference.

[00142] This was supplemented by an analysis based on Mahalanobis-distance (M-dist). M-dist is a multivariate measure of the distance between a single gene expression value from a patient and the mean of a pool of samples from controls. M-dist is expected to have a chi-square distribution with degrees of freedom equal to the number of genes. An arbitrary cut-off point, such as the 95th percentile, is chosen, below which most individual control values will fall. Thus an experimental subject with an M-dist sample value above this criterion can be thought of as being significantly different from a control sample.

[00143] M-dist values can be determined for either each individual biopsy or swab removed from an individual, or for the mean of gene expression values from all samples taken from an individual. These M-dist values can then be plotted on a graph, with the value from each sample or each individual represented by a single point. The sensitivity and specificity of the approach can be readily visualized from these plots. The sensitivity is the proportion of values in the experimental group that are above the 95th percentile--represented as a horizontal line on the graph--while the specificity is the proportion of all values above the line which belong to individuals in the experimental group.

[00144] Biopsies of colonic mucosa, from rectosigmoid or rectal areas, were taken from subjects during the course of colonoscopy. The subjects included individuals with adenomatous polyps, the precursor of most colon cancers; individuals with a family history or self history of cancer; and individuals with no polyps or family/self history, who served as normal controls. In all cases, the biopsies were composed of normal appearing mucosa.

[00145] In addition, mucosal samples were obtained from individuals in all these groups by a rectal smear, using a small

anoscope. A small brush was inserted through the anoscope several centimeters into rectum, and cells removed by gentle scraping.

[00146] Total RNA was extracted from each tissue sample, and reverse transcriptase used to convert RNA to cDNA. The expression of each of fifteen genes was then determined using PCR, with primers designed to amplify each gene.

[00147] Mahalanobis (M-dist) was selected as the measure of statistical significance because it summarizes in a single number the differences between a pattern of gene expression for any individual against the average of a pool of individuals, taking into account variability of each gene's expression and correlations among pairs of genes. This allowed us to determine on a probability scale, how different one gene expression pattern is from another. First, for each control biopsy, The M-dist was calculated from the multivariate mean of the other normal control biopsies. Then an M-dist was computed for each biopsy from each individual with polyps, family/self history of cancer, in which M-dist measured the individual's multivariate distance (i.e., difference in pattern of expression) from the pooled mean of the normal control samples. Using this approach, one can determine an upper bound for the normal controls, at any arbitrary level of significance, such as the 95th percentile. This allows analysis of significance of gene expression values of any individual experimental patient compared with the pool of normal controls.

[00148] Figure 1 shows the Mahalanobis distance for biopsy samples, taken from (left to right), controls, resected colon cancer, individuals with family history, and individuals with polyps. Each circle represents the M-distance of a single tissue sample, and all the circles in a single vertical line represent samples from a single individual. The horizontal line represents an M-dist corresponding to the 95th percentile for normal controls, so that any values above this line are significantly different from the pooled normal control values at a significance level of $p < 0.05$ (i.e. result is not like that for normal controls).

[00149] As expected, most of the samples from control individuals (99/104) fell below the 95th percentile. Four out of seventeen individuals had at least one sample above the line, and

just one 1/17 had two samples. In contrast, all biopsy samples from resected colon cancer tissue had M-dist values above the 95th percentile, and for 6/7 individuals, each value was far above the line ($p < 0.001$). For individuals with family history and individuals with polyps, some samples were above the 95th percentile and some below it, but all 13 individuals with family history had at least one sample above the line, as did 21/24 (87.5%) individuals with polyps. Ten of thirteen (77%) individuals with family history had more than one biopsy with an M-dist value above the line, while 14/24 (58%) individuals with polyps did.

[00150] Figure 1B shows analysis carried out on a second patient pool, one including individuals with no polyps or family/self history(Control), individuals with family history, individuals with polyps. The results are similar to those of the earlier study. All of the control biopsies had M-dist values below the 95th percentile. Fifteen of eighteen (83%) individuals with family history had at least one value above this percentile, while 4/9 (44%) individuals with polyps did.

[00151] Figure 1C shows the same analysis carried out on rectal smear samples taken from the same individuals used in the study presented in Figure 1B. All but one normal control biopsy were at or below the 95th percentile. 15/17 (88%) individuals with family/self history had at least one M-dist value above the 95th percentile, and 13/17 (76.5%) had at least two values above it. All 9 individuals with polyps had at least one value above the 95th percentile, and 5/9 (56%) had at least two values above this criterion. In addition, all smear taken from known colon cancer from two individuals had M-dist values far above the 95th percentile.

[00152] Figure 2A-B show the similar analysis based upon a swab. Figure 2A shows a 90 patient study of gene expression values for 16 genes from each subject, controls tend to fall below the 95% chi-square distribution line. A tendency of subjects with cancer fall above the line can be seen at the far right. Figure 2B shows the 95% chi-square distribution of gene analysis from buccal swabs of 21 controls and 8 cancer subjects. The data demonstrate that a buccal swab and analysis of a panel of genes in the sample can be

used to identify subject with a gene expression profile different than that a normal control. The difference being indicative of a risk factor for colorectal cancer.

[00153] Colon cancer is the result of a progression of molecular and cellular changes in the mucosal tissue lining the colon. While these changes are not completely understood, they are accompanied by alterations in the expression levels of many genes. Taking advantage of this fact, we have previously shown that normal appearing colon mucosa from individuals with polyps, family/self history of cancer has a different expression profile. The tissue samples from these studies were obtained by colonoscopy, but here we have shown that samples can also be obtained by rectal smear, a non-invasive procedure that can be carried out quickly and cheaply in any physician's office, without bowel preparation or anesthesia.

[00154] These results indicate that one can identify all cases of colon cancer and distinguish a high % of individuals with adenomatous polyps from those without polyps. Individuals at risk for cancer can be recommended for colonoscopies, while those with no risk may choose to avoid this costly and invasive procedure.

[00155] A number of embodiments have been described. Nevertheless, it will be understood that various modifications may be made without departing from the spirit and scope of the description. Accordingly, other embodiments are within the scope of the following claims.

What is claimed is

1. A method for determining if a subject has or is at risk of having a gastrointestinal disease or disorder comprising:

measuring an FHSH biomarker panel, a polyp biomarker panel, a rectal bleeding biomarker panel, a cancer biomarker panel or any combination thereof,

wherein a change in one or more of the biomarker panels relative to a control is indicative of a subject that has or is at risk of having a gastrointestinal disease or disorder.

2. The method of claim 1, comprising measuring an FHSH biomarker panel and comparing the measurements to a control wherein a change relative to the control is indicative that the subject has a predisposition or risk of developing a cancerous lesion.

3. The method of claim 1 or 2, wherein the FHSH biomarker panel is obtained by swab.

4. The method of claim 3, wherein the FHSH biomarker panel based upon a swab comprises one or more of the biomarkers Groα, CD44, and COX1.

5. The method of claim 1 or 4, wherein a subject has a family history or self history if the FHSH biomarker panel measurements show a change in expression of a biomarker of at least 15% compared to a control subject population.

6. The method of claim 2, wherein the FHSH biomarker panel is obtained by biopsy.

7. The method of claim 6, wherein the FHSH biomarker panel based upon a biopsy comprises one or more of the biomarkers GROα, PPARδ, SAA1, COX1 and CXCR2.

8. The method of claim 1, 6, or 7, wherein a subject has a family history or self history if the FHSH biomarker panel

measurements show a change in expression of a biomarker of at least 15% compared to a control subject population.

9. The method of claim 2, wherein a subject that is identified as having a predisposition or risk of developing a polyp or cancerous lesion is monitored for a polyp biomarker panel.

10. The method of claim 9, wherein the polyp biomarker panel is obtained by a swab.

11. The method of claim 10, wherein the polyp biomarker panel based upon a swab comprises one or more of the biomarkers CD44, PPARγ, and COX1.

12. The method of claim 1, 10 or 11, wherein a subject has a risk of having or has a polyp if the polyp biomarker panel measurements show a change in expression of a polyp biomarker of at least 15% compared to a control subject population.

13. The method of claim 9, wherein the polyp biomarker panel is obtained by a biopsy.

14. The method of 13, wherein the polyp biomarker panel based upon a biopsy comprises one or more of the biomarkers Groα, CXCR2, and PPARδ.

15. The method of claim 1, 13 or 14, wherein a subject has a risk of having or has a polyp if the polyp biomarker panel measurements show a change in expression of a polyp biomarker of at least 15% compared to a control subject population.

16. The method of claim 2 or 9, wherein the subject is further monitored for a cancer biomarker panel.

17. The method of claim 16, wherein the cancer biomarker panel is obtained by swab in the absence of an RNA protection cocktail.

18. The method of claim 17, wherein the cancer biomarker panel based upon a swab in the absence of an RNA protection cocktail comprises one or more of the biomarkers PPARα, CXCR2, cMyc and CD44.

19. The method of claim 1, 17, or 18, wherein a subject has a risk of having or has a colorectal cancer if the cancer biomarker panel measurements show a change in expression of a cancer biomarker of at least 15% compared to a control subject population.

20. The method of claim 16, wherein the cancer biomarker panel based upon a swab is obtained in the presence of an RNA protection cocktail.

21. The method of claim 20, wherein cancer biomarker panel based upon a swab in the presence of an RNA protection cocktail comprises one or more of the biomarkers COX2 and IL-8.

22. The method of claim 1, 20, or 21, wherein a subject has a risk of having or has a colorectal cancer if the cancer biomarker panel measurements show a change in expression of a cancer biomarker of at least 15% compared to a control subject population.

23. The method of claim 1, comprising meausuring a polyp or cancer biomarker panel and comparing the measurements to a control wherein a change relative to the control is indicative that the subject has or is at risk of developing a polyp or cancerous lesion.

24. The method of claim 23, wherein a subject that is identified as having a predisposition or risk of developing a polyp is monitored for a cancer biomarker panel.

25. The method of claim 23, wherein the polyp biomarker panel is obtained by a swab.

26. The method of claim 25, wherein the polyp biomarker panel based upon a swab comprises one or more of the biomarkers CD44, PPARγ, and COX1.

27. The method of claim 26, wherein a subject has a risk of having or has a polyp if the polyp biomarker panel measurements show a change in expression of a polyp biomarker of at least 15% compared to a control subject population.

28. The method of claim 23, wherein the polyp biomarker panel is obtained by a biopsy.

29. The method of 28, wherein the polyp biomarker panel based upon a biopsy comprises one or more of the biomarkers Groα, CXCR2, and PPARδ.

30. The method of claim 29, wherein a subject has a risk of having or has a polyp if the polyp biomarker panel measurements show a change in expression of a polyp biomarker of at least 15% compared to a control subject population.

31. The method of claim 23 or 24, wherein the cancer biomarker panel is obtained by swab in the absence of an RNA protection cocktail.

32. The method of claim 31, wherein the cancer biomarker panel based upon a swab in the absence of an RNA protection cocktail comprises one or more of the biomarkers PPARα, CXCR2, cMyc and CD44.

33. The method of claim 31, wherein a subject has a risk of having or has a colorectal cancer if the cancer biomarker panel measurements show a change in expression of a cancer biomarker of at least 15% compared to a control subject population.

34. The method of claim 23 or 24, wherein the cancer biomarker panel based upon a swab is obtained in the presence of an RNA protection cocktail.

35. The method of claim 34, wherein cancer biomarker panel based upon a swab in the presence of an RNA protection cocktail comprises one or more of the biomarkers COX2 and IL-8.

36. The method of claim 34, wherein a subject has a risk of having or has a colorectal cancer if the cancer biomarker panel measurements show a change in expression of a cancer biomarker of at least 15% compared to a control subject population.

37. A method of determining whether a subject has rectal bleeding comprising measuring a rectal bleeding biomarker panel, wherein a subject that is positive for the panel has rectal bleeding.

38. The method of claim 37, wherein the rectal bleeding biomarker panel is based upon a swab.

39. The method of claim 38, wherein the rectal bleeding biomarker panel comprises one or more of the biomarkers COX2, OPN, PPARγ, COX1 and GROα.

40. The method of claim 39, wherein a subject has a risk of having or has rectal bleeeding if the biomarker panel measurements show a change in expression of a rectal bleeding biomarker of at least 15% compared to a control subject population.

41. The method of claim 38, wherein a subject found to have a rectal bleeding biomarker panel has a non-cancerous inflammtory disease or disorder of the gastrointestinal tract.

42. The method of claim 37, wherein the rectal bleeding biomarker panel is based upon a biopsy.

43. The method of claim 42, wherein the rectal bleeding biomarker panel based upon a biopsy comprises one or more of the biomarkers Groα, Groγ, PPARδ and SAA1.

44. The method of claim 43, wherein a subject has a risk of having or has rectal bleeding if the rectal bleeding biomarker panel measurements show a change in expression of a rectal bleeding biomarker of at least 15% compared to a control subject population.

45. A kit for carrying out the method of claim 1, 8, or 37.

46. A kit for carrying out the method of claim 16.

47. A kit for carrying out detection of a FHSH biomarker panel, a polyp biomarker panel, a cancer biomarker panel, a rectal bleeding biomarker panel or any combination thereof.

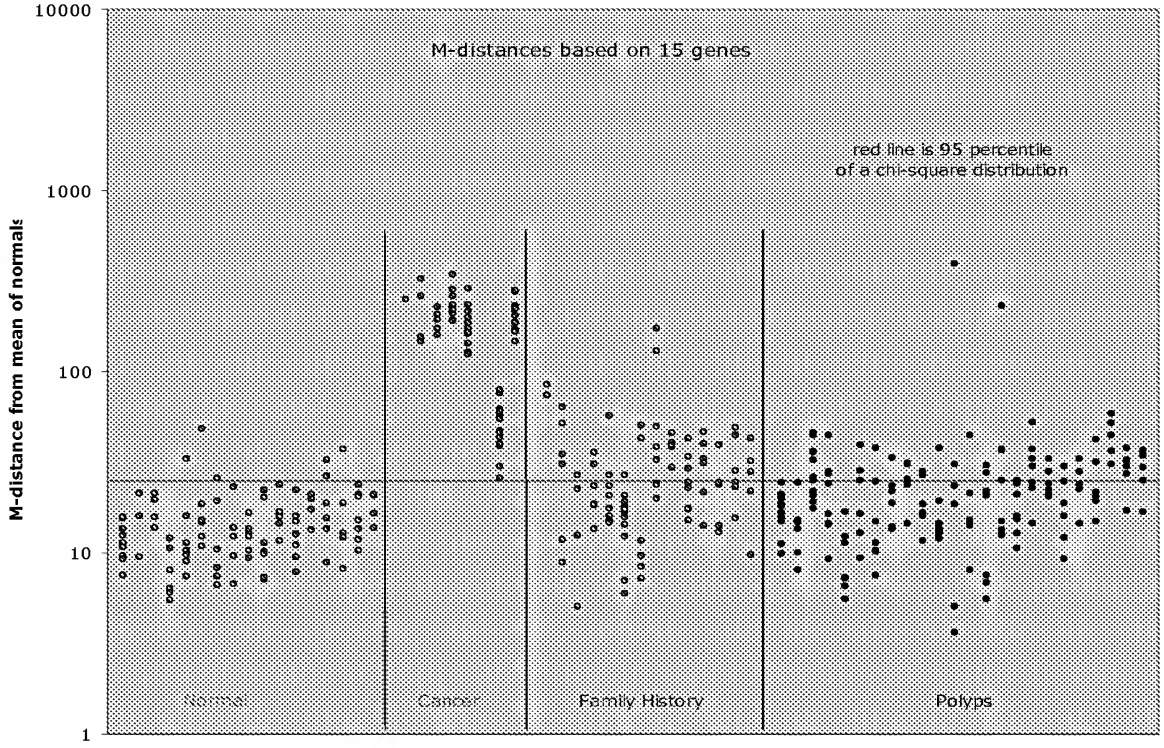

FIGURE 1A

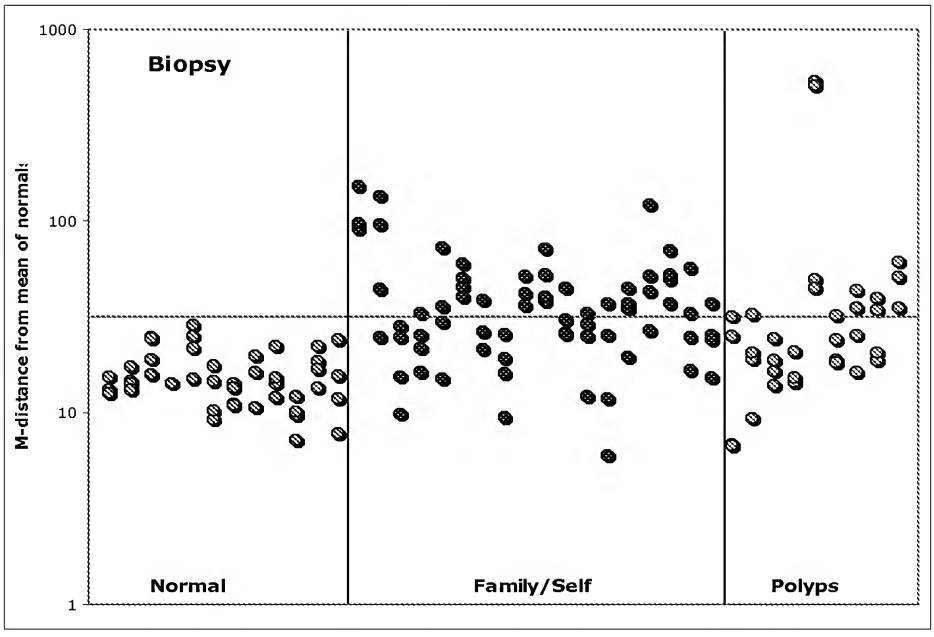

FIGURE 1B

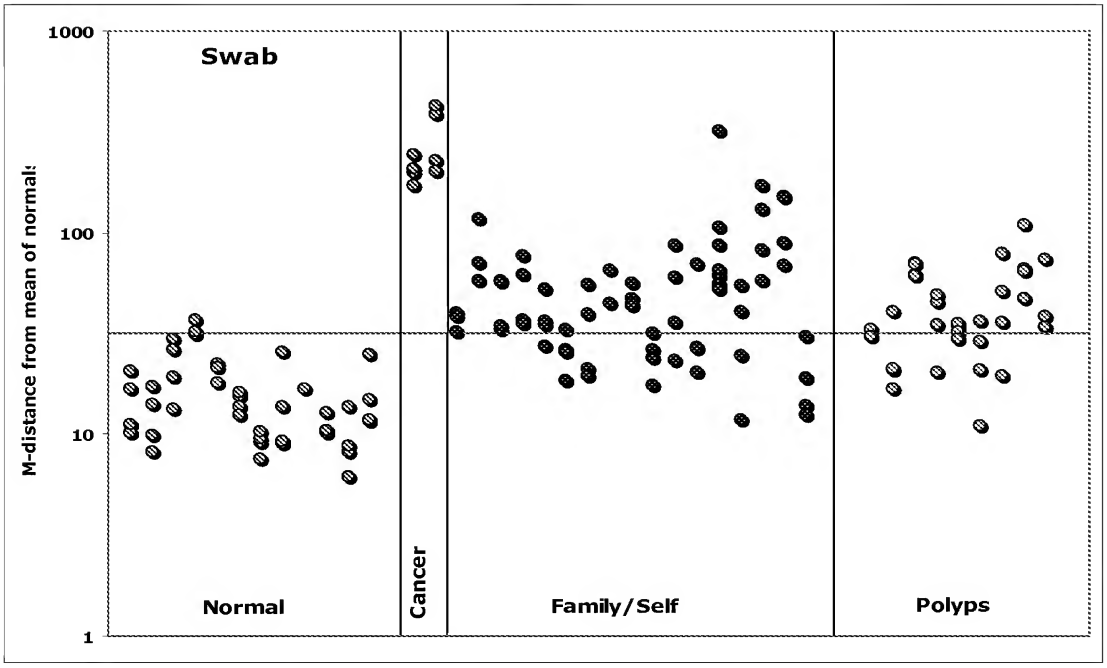
Swab
1000
100
10
1
M-distance from mean of normal
Normal
Cancer
Family/Self
Polyps

FIGURE 1C

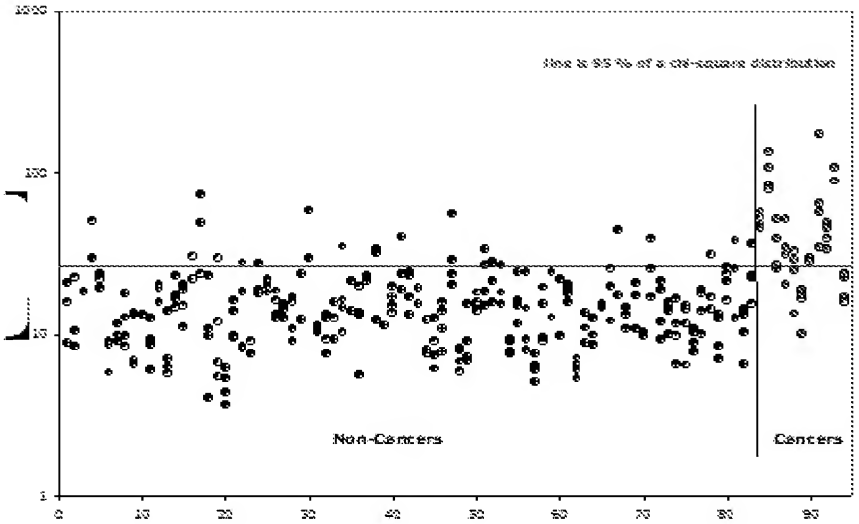

FIGURE 2A

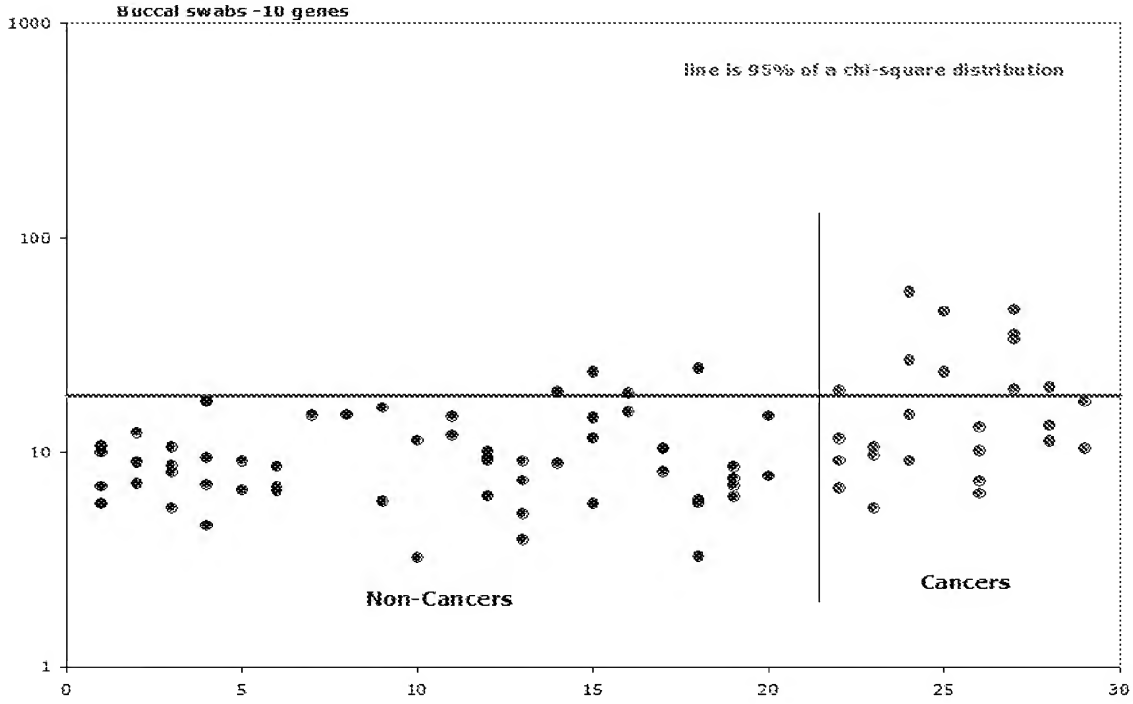

FIGURE 2B

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2008/071090

A. CLASSIFICATION OF SUBJECT MATTER

C12M 1/40(2006.01)i, C12Q 1/58(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 8: C12M 1/40, C12N 15/09, C12Q 1/00, C12Q 1/58, C12Q 1/68, G01N 33/567, G01N 33/574

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKIPASS(KIPO internal), USPTO, ScienceDirect (gastrointestinal, polyp, diagnostic, biomarker, control, and similar terms)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 20050014165 A1 (California Pacific Medical Center) 20 January 2005 See Abstract; Claims 1, 29, 65, 79-81.	46, 47
A	US 6423491 B1 (Howe, J.R.) 23 July 2002 See Abstract; Claims 1, 10, 30.	46, 47
A	JP 2005080524 A (Hidetoshi, K. et al.) 31 March 2005 See Abstract, Claims 1-3, 7-10, 29-38.	46, 47

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents,such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
02 DECEMBER 2008 (02.12.2008)	**02 DECEMBER 2008 (02.12.2008)**

Name and mailing address of the ISA/KR	Authorized officer
Korean Intellectual Property Office Government Complex-Daejeon, 139 Seonsa-ro, Seo-gu, Daejeon 302-701, Republic of Korea	JEONG, JAE CHEOL
Facsimile No. 82-42-472-7140	Telephone No. 82-42-481-8385

Form PCT/ISA/210 (second sheet) (July 2008)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2008/071090

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.: 1-44
because they relate to subject matter not required to be searched by this Authority, namely:

Claims 1-44 pertain to diagnostic methods, and thus relate to a subject matter which this International Searching Authority is not required, under Article 17(2)(a)(i) of the PCT and Rule 39.1(iv) of the Regulations under the PCT, to search.

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 19, 22, 45
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required addtional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2008/071090

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2005014165 A1	20.01.2005	AU 2004259431 A1	03.02.2005
		CA 2534633 A1	03.02.2005
		EP 1654526 A2	10.05.2006
		JP 2007512801A	24.05.2007
		KR 1020060034712 A	24.04.2006
		WO 2005010486 A2	03.02.2005
US 6423491 B1	23.07.2002	None	
JP 2005080524 A	31.03.2005	None	